The Cross-Sectional Distribution of Fund Skill Measures

Laurent Barras, Patrick Gagliardini, and Olivier Scaillet[*]

This version, October 18, 2018

ABSTRACT

We develop a simple, non-parametric approach for estimating the entire distribution of skill. Our approach avoids the challenge of correctly specifying the distribution, and allows for a joint analysis of multiple measures—a key requirement for examining skill. Our results show that more than 85% of the funds are skilled at detecting profitable trades, but unskilled at overriding capacity constraints. Aggregating both skill dimensions using the value added, we find that around 70% of the funds are able to generate profits. The average value added after funds have reached their long-term size equals 7.1 mio. per year, which represents two thirds of the optimal value predicted by neoclassical theory. For all skill measures, the distribution is highly non-normal and reveals a strong heterogeneity across funds.

I Introduction

Over the past 50 years, the academic literature on mutual funds has largely focused on mutual fund performance. For instance, Carhart (1997), Elton et al. (1993), and Jensen (1968) find that the aggregate alpha net of fees and trading costs is negative, while recent studies further show that the majority of funds deliver negative net alphas (e.g., Barras, Scaillet, and Wermers (2010), Harvey and Liu (2018a)). Far less attention has been devoted to the analysis of mutual fund skill.[1] Whereas these two notions are often used interchangeably, they differ in important ways. Skill is defined from the viewpoint of funds—it measures their ability to create value through their investment and trading decisions. Performance is defined from the viewpoint of investors—it measures whether the value created by the funds, if any, is passed on to them.

In this paper, we develop a novel approach for estimating the cross-sectional distribution of fund skill. Our approach is non-parametric—it does not impose any parametric structure on the shape of the distribution. As such, it brings two key advantages for examining skill. First, it avoids the challenge of correctly specifying the skill distribution. For one, theory offers little guidance to specify the mean, dispersion, or asymmetry of skill across funds. Data cannot be used to guide specification either—a simple histogram of the estimated skill measures is plagued by estimation errors and thus biased.

Second, our approach accommodates the analysis of multiple skill measures. This joint analysis is key because funds are skilled along several dimensions. In a world with decreasing returns to scale, Berk and Green (2004; BG hereafter) show that funds can be skilled (i) at detecting profitable trading ideas, or (ii) at mitigating the negative impact of capacity constraints. Furthermore, it is natural to aggregate both dimensions to infer the overall skill level of each fund. This step requires an aggregate skill measure such as the value added proposed by Berk and van Binsbergen (2015).

In addition to its flexibility, our novel approach is simple, widely applicable, and supported by econometric theory. It only involves simple manipulations of the fund skill measures estimated using time-series regressions. It provides a unified framework that is applicable to all the descriptive statistics of the skill distribution (density, cumulative function, quantiles, moments (e.g., mean, skewness)). Finally, it rests on solid statistical foundations that allow us to infer the asymptotic properties of the different estimators.

Our non-parametric approach departs from standard parametric/Bayesian approaches that require a full parametric specification of the distribution (e.g., Jones and Shanken

---

[1]Notable exceptions include Berk and van Binsbergen (2015), Grinblatt and Titman (1989), Pastor, Stambaugh, and Taylor (2015), and Wermers (2000).

1

(2005)). In the context of skill evaluation, these approaches are less appealing because they are prone to misspecification errors and cannot easily handle multiple skill measures. While a joint specification across all measures is subject to the curse of dimensionality, a separate specification is prone to inconsistencies because the skill measures are intertwined. Another limitation of these approaches is that they rely on sophisticated, computer-intensive estimation methods (e.g., Gibbs sampling).

The main challenge in estimating the skill distribution is to control for bias. Given that the *true* skill measures are unobservable, we can only use as inputs the estimated skill measures. This introduces an Error-in-Variable (EIV) bias that is reminiscent of the well-known EIV bias in the two-pass regression approach (Shanken (1992)). To address this issue, we derive a closed-form expression of the EIV bias that can be easily computed and interpreted. In essence, the bias adjustment consists of reducing the large tail probabilities obtained with the estimated skill measures. We further validate this procedure through an extensive Monte-Carlo analysis.

Our empirical analysis is based on monthly return data for all actively managed US equity funds between January 1979 and December 2015. We examine the skill distribution for the entire population, as well as different groups sorted on investment styles and fund characteristics. We begin by measuring the two skill dimensions using the BG model in which the fund gross alpha α_{it} depends on its lagged size $q_{i,t-1}$: $\alpha_{it} = a_i - b_i q_{i,t-1}$. The first skill dimension—the first dollar (fd) alpha a_i—determines the fund ability to identify profitable trades. The second dimension—the size coefficient b_i—captures the fund ability to resist capacity constraints. We then aggregate the two dimensions using the value added which determines the total profits earned by the fund from exploiting its skill. Our analysis explicitly distinguishes between two measures of the value added. The first one, defined as $va_i = E\left[\alpha_{i,t} q_{i,t-1}\right]$, determines the profits over the fund lifecycle. The second one, defined as $va_{i,ss} = E\left[\alpha_{i,t}\right] E\left[q_{i,t-1}\right]$, determines the steady state profits once the fund reaches its average size.

The ability of funds to detect profitable trades is both widespread and economically significant. The fd alpha a_i is positive for 87.5% of the funds and is equal to 3.5% per year, on average. In other words, most funds are able to detect undervalued stocks and correct for mispricing caused by noise trading. At the same time, only a handful of funds have the ability to override capacity constraints. The size coefficient b_i is positive for 85.9% of the funds and causes an average 1.5% decrease in annual alpha following a one standard deviation increase in size. Taken together, these fund-level results corrected for EIV bias provide strong support to the BG model in which investment skills and capacity constraints represent central features of the mutual fund industry.

2

The distributions of the two skill dimensions exhibit several common features. First, they reveal a substantial heterogeneity across funds. For instance, the volatility of the size coefficient is as large as its average (1.5%). Therefore, the standard approach of imposing a constant size coefficient ignores the diverse impact of capacity constraints across funds. Second, both distributions are highly non-normal with levels of skewness above 6.0. This implies that a minority of funds exhibit stellar investment skills— for 10% of the population, the fd alpha is above 7.8% per year. Third, adjusting for the EIV bias is essential for estimating the two skill distributions and their associated descriptive statistics. A simple analysis based on the estimated coefficients $\hat{a}_i$ and $\hat{b}_i$ is heavily biased—it largely overestimates the probability mass in the tails and completely fails to capture the asymmetry in skill (i.e., the unadjusted skewness is a mere 0.6).

Our joint analysis shows that the two skill dimensions are positively correlated. For instance, small cap funds exhibit both higher fd alphas and higher size coefficients than large cap funds. Similarly, high expense funds are skilled at detecting profitable trades, but unskilled at mitigating capacity constraints. These results help reconcile previous studies that interpret high expenses as a signal of either superior or inferior skill (e.g., Elton et al. (1993), Pastor, Stambaugh, and Taylor (2017)). They also imply that examining each dimension separately is not sufficient for determining the overall skill level. To this end, we need to aggregate both dimensions using the value added.

The analysis of the value added reveals that individual funds earn large profits from exploiting their skills. On average, the value added $va_{i,l}$ over the fund lifecycle is equal to 1.9 mio. per year. Once funds reach their average size, the steady-state value added is even larger as $va_{i,ss}$ equals 7.1 mio. per year, on average. Here again, controling for the EIV bias is essential. For one, the unadjusted proportion of funds with positive $va_{i,l}$ equals 49.5%, our bias-adjusted estimate is significantly higher at 64.1%. Similar to the two skill dimensions, the average value added provides a poor summary of the entire fund population because the distribution is highly volatile, positively skewed and fat-tailed. This large heterogeneity implies that some funds earn staggering profits, e.g., 10% of the funds produce a steady state value added $va_{i,ss}$ above 25.6 mio. per year.

Comparing the different fund groups, we find that the small cap funds earn larger profits than large cap funds. Therefore, their superior ability in detecting profitable trades compensates for their strong exposure to capacity constraints. We also observe significant differences between groups sorted on expense ratio and turnover—on average, low expense/turnover funds produce an incremental value added between 4.2 and 6.9 mio. per year. Therefore, these funds achieve better combinations of fd alphas and size coefficients and are rewarded by investors via larger money inflows.

Our non-parametric approach can also be used to study the distribution of the optimal value added va_i^* given by $\frac{a_i^2}{4b_i}$. This analysis is motivated by the neoclassical BG model which predicts that in a world where skilled funds are in scarce supply, they are able to earn the maximum profits va_i^*. On average, we find that the value added va_i represents only 9.5% of the optimal value va_i^*. This result may not be surprising because investors need time to learn about a_i and b_i (Pastor and Stambaugh (2012). Therefore, the size of the funds can be quite far from its optimal value. Past this learning phase, however, the empirical results are broadly supportive of the BG model—the steady state value added $va_{i,ss}$ represents, on average, 66.8% of the optimal value va_i^*.

Finally, we find that the gross alpha, which is widely used in previous work, contains limited information about skill. Our theoretical analysis shows that the gross alpha is aligned with the skill measures only if all funds choose specific compensation schemes. The empirical analysis of fund fees reveals that this is not the case because funds have a preference for managing a larger asset base. Therefore, the gross alpha becomes less informative—some funds can have superior skills and yet deliver lower gross alphas as they trade off lower fees for a larger size. Consistent with this result, we find that the fund gross alpha is weakly related to the different skill measures.

Our work is related to several strands of the literature. Recent papers use parametric/Bayesian approaches to infer the entire distribution of fund alphas (e.g., Chen, Cliff, and Zhao (2017), Jones and Shanken (2005), Harvey and Liu (2018a)) or their sensitivity to capacity constraints (Harvey and Liu (2018b)). Here, we apply a non-parametric approach to multiple skill measures. Several studies apply the False Discovery Rate approach to measure the proportions of funds with negative/positive performance (e.g., Avramov, Barras, and Kosowski (2013), Barras, Scaillet, and Wemers (2010), Ferson and Chen (2018)). This paper estimates the entire skill distribution (density, cumulative function, quantiles, moments) and thus encompasses the analysis of the proportions. Berk and van Binsbergen (2015) and Pastor, Stambaugh, and Taylor (2015) discuss the advantages of using the value added and the fd alpha. We build on these papers to estimate mutual fund skill. Finally, several studies provide evidence of capacity constraints at the aggregate level (e.g., Chen et al. (2004), Pastor, Stambaugh, and Taylor (2015)). Here, we examine the impact of capacity constraints at the individual fund level.

The remainder of the paper is as follows. Section II presents the different skill measures and describes our non-parametric approach. Section III describes the mutual fund data set and provides summary statistics. Section IV contains the empirical analysis of the different skill measures, and Section V concludes. The appendix contains the proofs of the different propositions, as well as additional empirical results.

II The Cross-Sectional Distribution of Skill Measures

A Measuring Fund Skill

A.1 Definition of the Skill Measures

The Two Dimensions of Skill We begin our presentation with the two dimensions of skill that funds potentiall exhibit, namely (i) their ability to generate profitable trading ideas, and (ii) their ability to mitigate the impact of capacity constraints. To this end, we build on the BG model and express the gross alpha of each fund i as a linear function of its lagged size $q_{i,t-1}$, i.e., $\alpha_{i,t} = a_i - b_i q_{i,t-1}$, for $i = 1, ..., n$, where n denotes the total number of funds.[2]

We measure the first skill dimension using the first dollar (fd) alpha a_i, which is defined as the fund abnormal return when its size equals zero.[3] This measure captures the quality of the fund investment ideas because it is not impacted by capacity contraints. In other words, the fd alpha is a "paper" return that is unencumbered by the drag of real world implementation (Perold and Salomon (1991)).

We measure the second skill dimension using the size coefficient b_i, which captures the fund sensitivity to capacity constraints. As discussed by BG, b_i is larger if the fund faces high execution costs for trading large orders (e.g., liquidity, price impact). A distinguishing feature of our specification is that b_i is fund-specific. This contrasts with the previous literature which commonly assumes a common coefficient for all funds (e.g., Chen et al. (2004), Pastor, Stambaugh, and Taylor (2015)). Capturing this heterogeneity is potentially important because there is a priori no reason why the impact of capacity constraints should be identical across funds.

To estimate the two skill dimensions, we use the following time-series regression:

$$r_{i,t} = a_i - b_i q_{i,t-1} + \beta'_i f_t + \varepsilon_{i,t}, \tag{1}$$

where $r_{i,t}$ is the fund gross excess return (before fees) over the riskfree rate, f_t is a K-vector of benchmark excess returns, and $\varepsilon_{i,t}$ is the error term.[4] To capture the heterogeneity across funds, we interpret Equation (1) as a random coefficient model in

---

[2] This linear specification is used, among others, by Berk and van Binsbergen (2015), Harvey and Liu (2018b), Pastor, Stambaugh, and Taylor (2015), and Pollet and Wilson (2008). In the appendix, we consider alternative specifications which leave our results largely unchanged.

[3] Here, we follow the terminology proposed by Pastor, Stambaugh, and Taylor (2015).

[4] In our baseline specification, we also use the four-factor model of Cremers, Petajisto, and Zitzewitz (2012) to estimate skill. To assess whether the model is correctly specified for mutual funds, we use the diagostic criterion by Gagliardini, Ossola, and Scaillet (2018). The results confirm that there is no remaining factor structure in the fund residuals $\varepsilon_{i,t}$ $(i = 1, ..., n)$.

which all the regression coefficients are random (e.g., Hsiao (2003)). In other words, we do not treat the coefficients a_i, b_i, and β_i as fixed parameters, but as random realizations from a continuum of funds. Under this sampling scheme, we can invoke cross-sectional limits in order to infer the common density function $\phi(a)$ and $\phi(b)$ from which the fd alpha and size coefficient of each fund are drawn.[5]

The Value Added To aggregate the two skill dimensions, we use the value added which determines the total profits earned by the fund through its investment and trading decisions. This measure has a powerful economic interpretation—it is similar to the rent earned by a monopolist obtained by multiplying the markup price of the good by the total quantities sold.

Formally, we consider two formulations of the value added. The first one, proposed by Berk and van Binsbergen (2015), measures the value added during the fund entire lifecycle (1):

$$va_{i,l} = \plim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \alpha_{i,t} \cdot q_{i,t-1} = E\left[\alpha_{i,t} q_{i,t-1}\right] = a_i q_{1,i} - b_i q_{2,i}, \qquad (2)$$

where $q_{i,1}$, $q_{i,2}$ are equal to the time-series averages of the fund size and its squared value, i.e., $q_{1,i} = E[q_{i,t-1}]$ and $q_{2,i} = E[q_{i,t-1}^2]$. In other words, va_i captures the average profits across different levels of fund size. The second one corresponds to the steady state (ss) value added by the fund once it reaches its average (steady state) size $q_{1,i}$:

$$va_{i,ss} = \plim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \alpha_{i,t} \cdot \plim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} q_{i,t-1} = E\left[\alpha_{i,t}\right] E\left[q_{i,t-1}\right] = a_i q_{1,i} - b_i q_{1,i}^2. \qquad (3)$$

The difference between the two measures is equal to the covariance between the gross alpha and size: $va_{i,l} - va_{i,ss} = cov(\alpha_{i,t}, q_{i,t-1}) = -b_i var(q_{i,t-1}) = -b_i(q_{2,i} - q_{1,i}^2)$. Therefore, $va_{i,ss}$ is always larger than $va_{i,l}$ as long as the fund is sensitive to capacity constraints ($b_i > 0$).

As shown in Equations (2)-(3), $va_{i,l}$ and $va_{i,ss}$ depend on the two two skill dimensions a_i and b_i. We can therefore build on the random coefficient model in Equation (1) to infer the common density functions $\phi(va_l)$ and $\phi(va_{ss})$ from which the two measures of the valued added are drawn.

A.2 The Relation between Skill and the Gross Alpha

The skill measures presented above differ from the fund gross alpha α_i, which is defined as the average abnormal return from the following regression: $r_{i,t} = \alpha_i + \beta_i' f_t + \varepsilon_{i,t}$. Whereas the gross alpha is commonly used in the literature, it may provide limited information about skill because it does not control for fund size.[6] Going back to the analogy with the monopolist, using this measure is akin to measuring the monopolist rent with the markup price of the goods, regardless of how much quantity is sold.

The exact relation between the gross alpha and the skill measures depends on how fund size is determined in equilibrium. A natural benchmark is the neoclassical BG model in which skilled managers maximize profit and rational investors compete for performance such that the gross alpha equals fees, i.e., $\alpha_i = f_{e,i}$. Following the notation of Berk and van Binsbergen (2015), we write the (benchmark-adjusted) fund total revenue and cost from active management as $r_i = a_i q_i$ and $c_i = b_i q_i^2$. Using the standard first order condition, we obtain the optimal size $q_i^* = \frac{a_i}{2b_i}$, and the optimal value added $va_i^* = r_i^* - c_i^* = a_i q_i^* - b_i q_i^{*2} = \frac{a_i^2}{4b_i}$.[7]

A key insight from this model is that fund fees do not change the value added: (i) if the fund chooses low fees, it receives additional money from investors ($q_i - q_i^* > 0$) which is passively indexed to keep $r_i^* - c_i^*$ unchanged; (ii) if the fund chooses high fees, it can sell the index short and invest the proceeds in the fund ($q_i^* - q_i > 0$) to obtain $r_i^* - c_i^*$. However, the choice of fees determine the equilibrium size q_i and thus the relations between the skill measures. To see this point, we examine four hypothetical compensation schemes in which managers set fees based on specific rules.

Scheme I (fd alpha). Funds set fees at $f_{e,i}^a$ such that they operate at the profit-mazimizing size q_i^*. We have $\alpha_i = f_{e,i}^a = \frac{r_i^* - c_i^*}{q_i^*} = \frac{a_i q_i^* - b_i q_i^{*2}}{q_i^*} \div a_i$. Therefore, the gross alpha captures the fd alpha (first skill dimension).

Scheme II (size coefficient). Funds set fees at $f_{e,i}^b$ such that they operate at the squared optimal size q_i^{*2}. We have $\alpha_i = f_{e,i}^b = \frac{r_i^* - c_i^*}{q_i^{*2}} = \frac{a_i q_i^* - b_i q_i^{*2}}{q_i^{*2}} = b_i$. Therefore, the gross alpha captures the size coefficient (the second skill dimension).

Scheme III (value added). Funds set fees at $f_{e,i}^{va}$ such that the size remains constant across all funds at $\bar{q}$. The gross alpha is given by $\alpha_i = f_{e,i}^{va} = \frac{r_i^* - c_i^*}{\bar{q}} = \frac{a_1 q_1^* - b_1 q_1^{*2}}{\bar{q}} =$

---

[6] The gross alpha is examined, among others, by Baks, Metrick, and Wachter (2001), Barras, Scaillet, and Wermers (2010), Carhart (1997), Grinblatt and Titman (1989), Jensen (1968), Jones and Shanken (2005), Wermers (2000).

[7] Our analysis largely builds on that of Berk and van Binsbergen (2015) which already discusses the relation between the gross alpha and the value added. Our contribution is to include all four skill measures (fd alpha, size coefficient, value added, gross alpha) and examine their relations under different compensation schemes.

$\bar{q}\frac{a_i^2}{4b_i} \div va_i^*$. Therefore, the gross alpha captures the value added.

Scheme IV (no information). Here, funds choose fees $f_{e,i}^{no}$ at levels that differ from $f_{e,i}^a$, $f_{e,i}^b$, and $f_{e,i}^{va}$.[8] We have $\alpha_i = f_{e,i}^{no} = \frac{r_i^*-c_i^*}{q_i^{no}} = \frac{a_1 q_1^* - b_1 q_1^{*2}}{q_i^{no}} = \frac{a_i^2}{4q_i^{no}b_i}$, where $q_i^{no}(f_{e,i}^{no}) = \frac{a_i^2}{4f_{e,i}^{no}b_i}$. Therefore, the gross alpha is uninformative about skill (i.e., $\alpha_i \neq a_i$, $\alpha_i \neq b_i$, $\alpha_i \neq va_i$). For instance, suppose that fund A is more skilled than fund B on every dimension, but chooses the same level of fees ($f_{e,A}^{no} = f_{e,B}^{no} = \bar{f}_e$). A comparison based on the gross alpha fails to capture any skill difference because we have $\alpha_A = \alpha_B = \bar{f}_e$.

Table I summarizes the above analysis and reveals that the four compensation schemes yield different predictions for fund size and fees. As show in Section IV.D, we can examine these predictions empirically to shed light on the information content of the gross alpha. Under Scheme I, we observe a moderate cross-sectional variation in fees and size as funds choose different fees to reach the fund-specific optimal size. Under Scheme II, the fees must be tiny so that funds grow large and reach the squared optimal size. Under Scheme III, the fund size is the same for all funds. Finally, Scheme IV predicts a large cross-sectional variation in size because it corrects for the arbitrary fees set by funds.

<div align="center">Please insert Table I here</div>

B   Overview of the Non-Parametric Approach

B.1   General Motivation

In this section, we describe the approach for estimating the cross-sectional distribution of fund skill. For simplicity, we denote the skill measure by m_i, where $m_i \in \{a_i, b_i, va_{i,l}, va_{i,ss}\}$ encompasses all four measures presented above. Our approach is non-parametric, i.e., it estimates the skill distribution without imposing any structure on its shape. As such, it provides several key advantages.

First, our approach is immune to misspecification errors. This is not the case for standard Bayesian/parametric approaches because they require to specify the shape of the distribution. In the context of skill, choosing the correct specification is challenging. For one, theory can be used to anchor the distribution of net alpha around zero, but offers no such guidance for skill. Whereas we can gain flexibility by specifying a mixture of normals (e.g., Chen, Cliff, and Zhao (2017), Harvey and Liu (2018a)), determining the number of mixture components is technically difficult.[9]

---

[8] For instance, funds may choose fees to mitigate moral hazard (Habib and Johnson (2016)). In this case, fees are set at the minimum level such that skilled managers exert effort. As fees are driven by the manager-specific probability of shirking, they may be quite different from $f_{e,i}^a$, $f_{e,i}^b$, and $f_{e,i}^{va}$.

[9] Due to their non-regularity, there are many technical challenges concerning inference problems on

Second, it allows for a joint analysis of all four skill measures, which is extremely challenging with Bayesian/parametric approaches. On the one hand, specifying each skill distribution separately is likely to generate inconsistencies because the skill measures are theoretically related as per Equations (1)-(3). On the other, jointly specifying all skill distributions involves the daunting task of estimating a multivariate distribution whose marginals are potentially mixtures of distributions.

Third, the implementation of the non-parametric approach is simple and fast. Intuitively, it is akin to computing an histogram using as inputs the estimated skill measure $\hat{m}_i$ of each fund. In contrast, Bayesian/parametric approaches require sophisticated and computer-intensive Gibbs sampling and Expectation Maximum (EM) methods (e.g., Chen, Cliff, and Zhao (2017), Harvey and Liu (2018a), Jones and Shanken (2005)).

Fourth, it provides a unified framework for estimating the different characterizations of the skill distribution: (i) the density function ϕ, (ii) the moments (e.g., mean, variance), (iii) the cumulative function $\Phi(x) = prob[m_i \leq x] = \int_{-\infty}^{x} \phi(m)dm$, and (iv) the distribution quantile $q(p) = \Phi^{-1}(p)$, where p denotes the probability level.[10]

Last but not least, it comes with a full-fledged econometric theory. For each estimators that characterize the skill distribution, we derive its asymptotic distribution as the number of funds n and the number of return observations T grow large (simultaneous double asymptotics with $n, T \to \infty$).

B.2   Non-Parametric Estimation

We now explain the main steps of our non-parametric approach. For sake of brevity, we describe the procedure for estimating the skill density ϕ and relegate to the appendix the formal treatment of the three remaining measures (moment, cdf, quantiles), as well as the proofs of the different propositions.

To begin, we estimate the regression coefficients in Equation (1) for each of the n funds in the population. The vector of coefficients $\hat{\gamma}_i = \left(\hat{a}_i, \hat{b}_i, \hat{\beta}_i' \right)'$ for fund i ($i = 1, ...n$) is computed as

$$\hat{\gamma}_i = \hat{Q}_{x,i}^{-1} \frac{1}{T_i} \sum_{t=1}^{T} I_{i,t} x_{i,t} r_{i,t}, \tag{4}$$

where $I_{i,t}$ is an indicator variable equal to one if $r_{i,t}$ is observable (and zero otherwise),

---

various aspects of finite mixture models (see the review paper of Chen (2017)). For example, the classical distribution theory of the log likelihood test statistic does not hold for testing for no mixture (homogeneity) against mixture alternatives (Ghosh and Sen (1985)).

[10]To lighten the notation, we do not subscript the density ϕ and the other statistical quantities by the skill measure.

$T_i = \sum_{t=1}^{T} I_{i,t}$ is the total number of return observations for fund i, $x_{i,t} = [1, -q_{i,t-1}, f'_t]'$, is the vector of explanatory variables, and $\hat{Q}_{x,i} = \frac{1}{T_i} \sum_{t=1}^{T} I_{i,t} x_{i,t} x'_{i,t}$ is the estimated matrix of the second moments of $x_{i,t}$. Using the estimated coefficients along with the size and squared size time-series averages, $\overline{q}_{1,i} = \frac{1}{T_i} \sum_{t=1}^{T} I_{i,t} q_{i,t-1}$, $\overline{q}_{2,i} = \frac{1}{T_i} \sum_{t=1}^{T} I_{i,t} q_{i,t-1}^2$, we can then compute each of the four skill measures as

$$
\begin{aligned}
\text{Fd alpha} \quad &: \quad \hat{m}_i = \hat{a}_i, \\
\text{Size coefficient} \quad &: \quad \hat{m}_i = \hat{b}_i, \\
\underset{\text{(lifecycle)}}{\text{Value added}} \quad &: \quad \hat{m}_i = \widehat{va}_{i,l} = \hat{a}_i \overline{q}_{1,i} - \hat{b}_i \overline{q}_{2,i}, \\
\underset{\text{(steady state)}}{\text{Value added}} \quad &: \quad \hat{m}_i = \widehat{va}_{i,ss} = \hat{a}_i \overline{q}_{1,i} - \hat{b}_i \overline{q}_{1,i}^2.
\end{aligned} \tag{5}
$$

Given the unbalanced nature of the mutual fund panel, T_i can be small for some funds. As a result, the inversion of the matrix $\hat{Q}_{x,i}$ can be numerically unstable and yield an unreliable estimate $\hat{m}_i$. To address this issue, we follow Gagliardini, Ossola, and Scaillet (2016) and introduce a formal fund selection rule $\mathbf{1}_i^{\chi}$ equal to one if the following two conditions are met (and zero otherwise):

$$
\mathbf{1}_i^{\chi} = \mathbf{1} \left\{ CN \left(\hat{Q}_{x,i} \right) \leq \chi_{1,T}, \tau_{i,T} \leq \chi_{2,T} \right\}, \tag{6}
$$

where $CN \left(\hat{Q}_{x,i} \right) = \sqrt{eig_{max} \left(\hat{Q}_{x,i} \right) / eig_{\min} \left(\hat{Q}_{x,i} \right)}$ is the condition number of the matrix $\hat{Q}_{x,i}$ defined as the ratio of the largest to smallest eigenvalues eig_{max} and eig_{min}, $\tau_{i,T} = T/T_i$ is the inverse of relative sample size T_i/T, and $\chi_{1,T}$, $\chi_{2,T}$ denote the two threshold values. The first condition $\{ CN \left(\hat{Q}_{x,i} \right) \leq \chi_{1,T} \}$ excludes funds for which the time series regression is poorly conditioned, i.e., a large value of $CN \left(\hat{Q}_{x,i} \right)$ indicates multicollinearity problems (Belsley, Kuh, and Welsch (2004), Greene (2008)). The second condition $\{ \tau_{i,T} \leq \chi_{2,T} \}$ excludes funds for which the sample size is too small. Both thresholds $\chi_{1,T}$ and $\chi_{2,T}$ increase with the sample size T—with more return observations, the fund coefficients are estimated with greater accuracy which allows for a less stringent selection rule.

Next, we estimate the skill density function using a standard non-parametric approach based on kernel smoothing.[11] The estimated density $\hat{\phi}$ at a given point m is

---

[11] See, for instance, Ait-Sahalia (1996), Ait-Sahalia and Lo (1998), and Stanton (1997) for applications of kernel density estimation in finance.

computed as

$$\hat{\phi}(m) = \frac{1}{nh} \sum_{i=1}^{n} \mathbf{1}_i^\chi K\left(\frac{\hat{m}_i - m}{h}\right), \tag{7}$$

where h is the vanishing smoothing bandwidth—similar to the length of histogram bars, the bandwidth h determines how many observations around point m we use for estimation. The function K is a symmetric kernel function that integrates to one. Because the choice of K is not a crucial aspect of nonparametric analysis, we use the standard Gaussian kernel $K(u) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{u^2}{2})$ for simplicity (see Silverman (1986)).

The following proposition examines the asymptotic properties of $\hat{\phi}(m)$ as the size of the fund population n and the number of return observations T grow large for a vanishing bandwidth h.

Proposition II.1 *As $n, T \to \infty$ and $h \to 0$ such that $nh \to \infty$ and $\sqrt{nh}(h^2 T + (1/T)^{\frac{3}{2}}) \to 0$, we have*

$$\sqrt{nh}\left(\hat{\phi}(m) - \phi(m) - bs(m)\right) \Rightarrow N\left(0, K_1 \phi(m)\right), \tag{8}$$

and the bias term $bs(m)$ is the sum of two components,

$$bs_1(m) = \frac{1}{2} h^2 K_2 \phi^{(2)}(m), \tag{9}$$

$$bs_2(m) = \frac{1}{2T} \psi^{(2)}(m), \tag{10}$$

where $K_1 = \int K(u)^2 du$, $K_2 = \int u^2 K(u) du$, $\phi^{(2)}(m)$ is the second derivative of the density $\phi(m)$ and $\psi^{(2)}(m)$ is the second derivative of the function $\psi(m) = \omega(m)\phi(m)$ with $\omega(m) = E(S_i | m_i = m)$. The term S_i is the asymptotic variance of the estimated centered skill measure $\sqrt{T}(\hat{m}_i - m_i)$ equal to $\plim_{T \to \infty}\left(\frac{\tau_{i,T}^2}{T} \sum_{t,s=1}^{T} I_{i,t} I_{i,s} u_{i,t} u_{i,s}\right)$. For each skill measure, the term $u_{i,t}$ is given by

$$
\begin{aligned}
\text{Fd alpha} \quad &: \quad u_{i,t} = e_1' Q_{x,i}^{-1} x_{i,t} \varepsilon_{i,t}, \\
\text{Size coefficient} \quad &: \quad u_{i,t} = e_2' Q_{x,i}^{-1} x_{i,t} \varepsilon_{i,t}, \\
\begin{matrix}\text{Value added}\\ \text{\scriptsize(lifecycle)}\end{matrix} \quad &: \quad u_{i,t} = q_{1,i} e_1' Q_{x,i}^{-1} x_{i,t} \varepsilon_{i,t} + a_i(q_{i,t-1} - q_{1,i}) \\
&\qquad - q_{2,i} e_2' Q_{x,i}^{-1} x_{i,t} \varepsilon_{i,t} - b_i(q_{i,t-1}^2 - q_{2,i}), \\
\begin{matrix}\text{Value added}\\ \text{\scriptsize(steady state)}\end{matrix} \quad &: \quad u_{i,t} = q_{1,i} e_1' Q_{x,i}^{-1} x_{i,t} \varepsilon_{i,t} + a_i(q_{i,t-1} - q_{1,i}) \\
&\qquad - q_{1,i}^2 e_2' Q_{x,i}^{-1} x_{i,t} \varepsilon_{i,t} - b_i 2 q_{1,i}(q_{i,t-1} - q_{1,i}), \tag{11}
\end{aligned}
$$

11

where e_1 (e_2) is a vector with one in the first (second) position and zeros elsewhere and $Q_{x,i} = E[x_{i,t}x'_{i,t}]$. Under a Gaussian kernel, the two constants K_1 and K_2 are equal to $\frac{1}{2\sqrt{\pi}}$ and 1, respectively.

Proof. See the appendix. ∎

Proposition II.1 yields several important insights. First, it shows that the estimated density function $\hat{\phi}(m)$ is asymptotically normally distributed, which facilitates the construction of confidence intervals. As shown in Equation (9), the width of this interval depends on the variance term $K_1\phi(m)$ which is higher in the peak of the density.[12]

Second, $\hat{\phi}(m)$ is a biased estimator of $\phi(m)$. Therefore, we can improve the density estimation by adjusting for the bias term $bs(m)$. Equations (10)-(11) reveal that $bs(m)$ has two distinct components. The first component bs_1 is the smoothing bias, which is standard in non-parametric density estimation (e.g., Silverman (1986), Wand and Jones (1995)). The second component bs_2, which is referred to as the error-in-variable (EIV) bias, is non-standard—it arises because we estimate ϕ using the estimated skill measure instead of the *true* one (i.e., $\hat{m}_i$ instead of m_i).

Finally, Proposition II.1 provides guidelines for the choice of the bandwidth. We show in the appendix that the choice of the optimal bandwidth h^*—the one that minimizes the Asymptotic Mean Integrated Squared Error (AMISE) of $\hat{\phi}(m)$—depends on the relationship between T and n: (i) if T is small relative to n ($n^{2/5}/T \to \infty$), h^* is proportional to $(nT)^{-\frac{1}{3}}$; (ii) if T is not small relative to n ($n^{2/5}/T \to 0$), h^* is proportional to $n^{-\frac{1}{5}}$.[13] Our Monte-Carlo analysis reveals that given our actual sample size, the two bandwidth choices produce similar results with a slight advantage to the first case. Motivated by these results, we use the following bandwidth in our baseline specification:

$$h^* = \left(\frac{K_2}{K_1} \int \phi^{(2)}(m)\psi^{(2)}(m)dm \right)^{-\frac{1}{3}} (n/T)^{-\frac{1}{3}}. \tag{12}$$

B.3 Adjusting for the Bias

Building on the insights of Proposition II.1, we show how to compute the bias-adjusted density estimator. Our approach consists of estimating the two bias terms $bs_1(m)$, $bs_1(m)$, and the optimal bandwidth h^* using a Gaussian reference model in which the fund skill measure m_i and the log of the asymptotic variance $s_i = \log(S_i)$ are

---

[12] Similar to Equation (7), Okui and Yanagi (2018) consider a kernel estimator for the density of the mean and autocorrelation of random variables. However, their distributional results differ from those derived in our regression context aimed at measuring fund skill.

[13] The AMISE is defined as the integrated sum of the leading terms of the asymptotic variance and squared bias of the estimated density $\hat{\phi}(m)$.

drawn from a bivariate normal distribution: $m_i \sim N(\mu_m, \sigma_m)$, $s_i \sim N(\mu_s, \sigma_s)$, and $corr(m_i, s_i) = \rho$.[14]

This simple reference model has several appealing properties. First, the estimators of the bias and the bandwidth are simple to compute because they are all available in closed form. Second, they are precisely estimated because they only depend on the five parameters of the normal distribution. Finally, we can perform a comparative static analysis on the two bias components, i.e., we can determine how their shapes change with the different parameters and gey some useful economic insights (as shown below).

These benefits are not shared by the alternative approach in which the bias terms are directly inferred from Equations (9)-(10) via a non-parametric estimation of the second-order derivatives $\phi^{(2)}$ and $\psi^{(2)}$. Estimating these derivative terms is notoriously difficult and generally leads to large estimation errors (e.g., Wand and Jones (1995; Section 2.12)).[15]

The following proposition provides the closed-form expressions for the two bias components and the optimal bandwidth as the size of the fund population n and the number of return observations T grow large for a vanishing bandwidth h.

Proposition II.2 *As $n, T \to \infty$ and $h \to 0$ such that $nh \to \infty$ and $\sqrt{nh}(h^2 T + (1/T)^{\frac{3}{2}}) \to 0$, the two bias components under the reference model are equal to*

$$bs_1^r(m) = \left[\frac{1}{2} K_2 h^2 \frac{1}{\sigma_m^2} (\bar{m}_1^2 - 1) \right] \frac{1}{\sigma_m} \varphi(\bar{m}_1), \tag{13}$$

$$bs_2^r(m) = \left[\frac{1}{2T} \exp\left(\mu_s + \frac{1}{2}\sigma_s^2 \right) \frac{1}{\sigma_m^2} (\bar{m}_2^2 - 1) \right] \frac{1}{\sigma_m} \varphi(\bar{m}_2), \tag{14}$$

where $\bar{m}_1 = \dfrac{m - \mu_m}{\sigma_m}$, $\bar{m}_2 = \dfrac{m - \mu_m - \rho\sigma_m\sigma_s}{\sigma_m}$, $\varphi(x) = \dfrac{1}{\sqrt{2\pi}} \exp(-\dfrac{1}{2}x^2)$ is the density of the standard normal distribution. In addition, the optimal bandwidth h^* is given by

$$h^* = \left[\frac{K_2}{K_1 2\sqrt{\pi}} \frac{3}{4\sigma_m^5} \left(\frac{\rho^4 \sigma_s^4}{12} - \rho^2 \sigma_s^2 + 1 \right) \exp\left(\mu_s + \frac{1}{2}\sigma_s^2(1 - \frac{\rho^2}{2}) \right) \right]^{-\frac{1}{3}} (nT)^{-\frac{1}{3}}. \tag{15}$$

Proof. See the appendix ∎

Building on Proposition II.2, we compute the bias-adjusted density $\hat{\phi}^*(m)$ using the following steps. First, we estimate the moments of the bivariate normal distribution in the reference model using the estimated quantities $\hat{m}_i$ and $\hat{s}_i$ ($i = 1, ..., n$). To compute $\hat{s}_i = \log(\hat{S}_i)$, we use the standard variance estimator of Newey and West (1987):

$$\hat{S}_i = \frac{\tau_{i,T}^2}{T} \sum_{t=1}^{T} I_{i,t} \hat{u}_{i,t}^2 + 2 \sum_{l=1}^{L} \left(1 - \frac{l}{L+1}\right) \left[\frac{\tau_{i,T}^2}{T} \sum_{t=1}^{T-l} I_{i,t} I_{i,t+l} \hat{u}_{i,t} \hat{u}_{i,t+l}\right], \qquad (16)$$

where $\hat{u}_{i,t}$ is obtained by plugging the estimated quantities for the chosen skill measure given in Proposition II.1, and L is the number of lags chosen to account for potential serial correlation. Second, we insert these estimated moments in Equations (13)-(15) to compute the bias terms $\widehat{bs}_1^r(m)$, $\widehat{bs}_2^r(m)$, and the optimal bandwidth h^*. Third, we remove the bias terms from the unadjusted density in Equation (7) to obtain the bias-adjusted density estimator

$$\hat{\phi}^*(m) = \hat{\phi}(m) - \widehat{bs}_1^r(m) - \widehat{bs}_2^r(m). \qquad (17)$$

While our approach allows for a simple bias adjustment, an important question is whether the estimated bias terms are sufficiently accurate when the data do not follow our simple reference model. To address this issue, we perform an extensive Monte-Carlo analysis that replicates the salient features of the data. The results summarized in the appendix reveal that the bias-adjusted estimator captures the *true* distribution $\phi(m)$ with remarkable accuracy.[16]

C   Analysis of the Density Bias

C.1   The Shape of the Two Bias Components

To shed light on the bias adjustment mechanism, we now study the shape of the two bias components in Equations (13)-(14). A key feature of the smoothing bias $bs_1^r(m)$ is that it depends on the total number of funds n. As n increases, h^* shrinks towards zero, which reduces the magnitude of $bs_1^r(m)$. With a population of several thousand funds, the contribution of $bs_1^r(m)$ thus becomes negligible. In contrast, the EIV bias $bs_2^r(m)$ depends on the number of observations T because it arises from the gap between $\hat{m}_i$ and m_i. Therefore, it can remain significant even if the fund population is large.

---

[16] Our results resonate with those reported by Silverman (1986) for the standard non-parametric density estimation without the EIV problem. He shows that the rule of thumb for the bandwidth choice, which relies on a Gaussian reference model, is quite robust to departures from normality.

The sign of the EIV bias is driven by the term $(\bar{m}_2^2 - 1)$ which is negative when m is close to the average μ_m and positive otherwise. As a result, $bs_2^r(m)$ is negative in the center of the distribution and positive in the tails. The intuition for this result is that the estimated skill measure is a noisy version of the *true* one ($\hat{m}_i = m_i +$ estimation noise). Therefore, the density obtained with $\hat{m}_i$ overestimates the probability of observing extreme skill levels, i.e., the unadjusted density $\hat{\phi}(m)$ is too flat.

Importantly, the shape of the EIV bias obtained with the reference model is representative of a large set of distributions. As shown in Proposition II.1, the *true* bias $bs_2(m)$ is a function of the second-order derivative of the *true* density $\phi^{(2)}(m)$. As long as this density peaks around its mean, $\phi^{(2)}(m)$ takes negative (positive) values in the center (tails) of the distribution—just like the bias obtained with the reference model.[17]

To illustrate, Figure 1 plots the two bias components $bs_1^r(m)$ and $bs_2^r(m)$ for the fd alpha obtained with our entire sample of funds. Consistent with our analysis, we find that the smoothing bias is close to zero because of the large population size ($n = 2,265$), while the EIV bias is significantly higher. If we integrate $bs_2^r(m)$ in the center of the distribution where its value is positive, we obtain a probability mass of 15%.

Please insert Figure 1 here

C.2 Comparative Statics for the EIV Bias

We can also use the reference model to perform a comparative static analysis on the EIV bias. There are three key parameters that determine the shape of $bs_2^r(m)$: (i) the volatility σ_m of the skill measure, (ii) the average level of the estimation variance $\mu_S = \exp\left(\mu_s + \frac{1}{2}\sigma_s^2\right)$, and (iii) the correlation ρ between the skill measure and the estimation variance.

A higher value for σ_m reduces the magnitude of $bs_2^r(m)$ because it makes the cross-sectional variation in the estimated measure more aligned with that of the *true* one (i.e., the relative importance of m_i over noise increases). A higher value for μ_S increases the magnitude of $bs_2^r(m)$ because the estimated skill measure becomes more volatile (i.e., the relative importance of noise over m_i increases). Finally, a higher value for ρ moves $bs_2^r(m)$ to the right because funds with higher skill are also more likely to exhibit higher estimation variance.

---

[17] The only case where $bs_2(m)$ differs from $bs_2^r(m)$ is if the *true* density $\phi(m)$ is a mixture of distributions whose components have means extremely far away from one another such that we have a trough instead of a peak at the mean of $\phi(m)$.

15

D   Extensions

D.1   Time-Varying Skill

We conclude our presentation with several useful extensions. We can use our non-parametric approach to examine the time-variation in individual fund returns. This analysis provides an extension of recent studies that examine how the gross alpha at the aggregate level varies with size, turnover, and business cycle conditions (e.g., Chen et al. (2004), Kacperczyk, van Nieuwerburgh, and Veldkamp (2014), Pastor, Stambaugh, and Taylor (2015, 2017)). To estimate the density $\phi(b_q)$ associated with each alpha predictor q $(q = 1, ..., Q)$, we generalize Equation (1) as: $a_i + \sum_{q=1}^{Q} b_{i,l} z_{q,i,t-1} + \beta_i' f_t + \varepsilon_{i,t}$, where $z_{l,i,t-1}$ is predictor q and $b_{i,l}$ denotes its impact on the fund gross alpha. Then, we replace $\hat{m}_i$ with the estimated coefficient $\hat{b}_{i,q}$ and modify $u_{i,t}$ accordingly.

D.2   Net Alpha

We can also evaluate mutual fund performance using the net alpha α_i^n, which measures the average abnormal return earned by investors. To estimate the cross-sectional density $\phi(\alpha^n)$, we estimate the net alpha of each fund using the following regression: $r_{i,t}^n = r_{i,t} - fe_{i,t} = \alpha_i^n + \beta_i' f_t + \varepsilon_{i,t}$, where $r_{i,t}^n$ denotes the fund net excess return. Then, we replace $\hat{m}_i$ with the estimated net alpha $\hat{\alpha}_i^n$ and modify $u_{i,t}$ accordingly.

D.3   Factor Beta

Finally, we can study the fund betas on the different risk factors. To compute the distribution $\phi(\beta_k)$ associated with each risk factor k $(k = 1, ..., K)$, we replace $\hat{m}_i$ with the estimated beta $\hat{\beta}_{i,k}$ and to modify $u_{i,t}$ accordingly. Our approach departs from previous work in which the statistical inference is typically based on the estimated betas without the EIV bias adjustement.

III   Data Description

A   Mutual Fund Data

We conduct our analysis on the entire population of actively managed US equity funds. We collect monthly data on net returns and size, as well as annual data on fees, turnover, and investment objectives from the CRSP database between January 1979 and December

2015 (444 observations).[18]. We measure the monthly gross return $r_{i,t}$ as the sum of the fund monthly net return and fees. The net return $r_{i,t}^n$ is computed as a value-weighted average of the net returns across all shareclasses using their beginning-of-month total net asset values. The monthly fees $fe_{i,t}$ are defined as the value-weighted average of the most recently reported annual fees across shareclasses divided by 12. We measure the fund size $q_{i,t-1}$, by taking the sum of the beginning-of-month net asset values across all shareclasses. Following Berk and van Binsbergen (2015), we adjust size for inflation by expressing all numbers in January 1, 2000 dollars.

Our initial sample includes all funds with a valid equity investment objective from Lipper, Strategic Insight, Wiesenberger, or CRSP. We exclude index funds and eliminate return observations of tiny funds by imposing a minimum size of \$15 million (see Chen et al. (2004), Pastor, Stambaugh, and Taylor (2015)). To apply the fund selection rules in Equation (6), we follow Gagliardini, Ossola, and Scaillet (2016) and select funds for which the condition number (CN) of the matrix $\hat{Q}_{x,i}$ is below 15 and the number of monthly observations T_i is above 60.[19] These selection criteria produce a final universe of more than 2,000 funds.[20]

We further group funds according to their investment styles (small cap, large cap, growth, and value). At the start of each month, we classify each fund using the style information provided by Lipper. If this information is missing, we use the investment objectives reported by Strategic Insight, Wiesenberger, and CRSP (see the appendix for additional details). A fund is included in a given group if its style corresponds to that of the group for a sufficiently long period such that the two selection rules are satisfied (i.e., $CN_i \leq 15$ and $T_i \geq 60$). We follow the same procedure to form terciles of funds based on their characteristics (expense ratios and turnover), where the monthly turnover is defined as the most recently observed ratio of min(buys,sells) on fund size (see Pastor, Stambaugh, and Taylor (2017)).

B Benchmark Models

To estimate the skill measures in Equations (1)-(3), we primarily use the benchmark model of Cremers, Petajisto, and Zitzewitz (2012; CPZ hereafter). Similar to the model of Carhart (1997), the CPZ model includes four factors, i.e., we have $f_t =$

---

[18]The starting date corresponds to the first month when the risk factors are available.

[19]The monthly returns need not be contiguous. We delete the observation following any missing returns because CRSP reports the cumulated return since the last reported observation.

[20]We compute $\hat{s}_i$ using a lag of three months ($L = 3$). To mitigate the impact of outliers on the estimated parameters in the reference model, we also exclude the values for $\hat{m}_i$ and $\hat{s}_i$ that are five standard deviation away from the mean.

$[r_{m,t}, r_{smb,t}, r_{hml,t}, r_{mom,t}]'$, where $r_{m,t}$, $r_{smb,t}$, $r_{hml,t}$, and $r_{mom,t}$ capture the excess returns of the market, size, value, and momentum factors. However, the CPZ model departs from the Carhart model in two respects: (i) $r_{m,t}$ is proxied by the excess return of the S&P500 index (instead of the CRSP market index), and (ii) the size and value factors are index-based, i.e., $r_{smb,t}$ is measured as the return difference between the Russell 2000 and the S&P500, while $r_{hml,t}$ is measured as the return difference between the Russell 3000 Value and the Russell 3000 Growth.

It is common for US equity managers to use as benchmarks the S&P500 and the Russell 2000, which both cover about 85% of the total market capitalization. As noted by CPZ, the Carhart model fails to price these passive, well-diversified indices—for instance, the Russell 2000 produces a Carhart alpha of -2.4% per year over the period 1980-2005. This implies that small-cap managers are likely be classified as unskilled if they use the Russell 2000 as benchmark. Motivated by these results, we use the CPZ model in our baseline specification, and later repeat the analysis with the Carhart model.

C   Summary Statistics

Table II reports summary statistics for value weighted portfolios of funds for the entire population and the different groups. In Panel A, we report the first four moments of the portfolio gross excess returns. In the entire population, the portfolio achieves a risk-return tradeoff similar to that of the aggregate stock market with a mean and volatility equal to 7.7% and 15.1% per year. It also exhibits a negative skewness (-0.8) and a positive kurtosis (5.4). The results are similar across groups except for small cap funds which produce higher mean and volatility.

In Panel B, we repeat the analysis for the estimated betas on the four factors in the CPZ model. Consistent with intuition, small cap funds are heavily exposed to the size factor (0.81). Whereas growth funds are negatively exposed to the value factor (-0.28), the opposite holds for value funds (0.20). We also find that high expense and high turnover funds tilt toward small cap, growth stocks.

Please insert Table II here

18

IV   Empirical Results

A   The Two Dimensions of Skill

A.1   The First Dollar Alpha

We begin our empirical analysis with the fd alpha which measures the fund ability to detect profitable trades. After estimating a_i for each fund using Equation (1), we apply our non-parametric approach to compute the cross-sectional density $\hat{\phi}^*(a)$ given in Equation (17). To describe the properties of this distribution, we also compute the bias-adjusted estimates of the (i) moments (mean, variance, skewness, kurtosis), (ii) proportions of funds with negative and positive fd alphas, $\hat{\pi}_a^-$ and $\hat{\pi}_a^+$, and (iii) distribution quantiles at 10% and 90%, $\hat{q}(10)$ and $\hat{q}(90)$.

Table III reveals that the fund investment skills are both economically large and widespread—on average, the fd alpha equals 3.5% per year and around 90% of the funds in the population are skilled. These results resonate with the calibration exercise in BG that yields a value for $\hat{\pi}_a^+$ around 80%. This implies that most funds are able to detect undervalued stocks based on their private information, and to correct for the mispricing caused by noise traders (e.g., Stambaugh (2014)).

Another striking feature of the fund population is the wide heterogeneity in skill, as captured by the 8.2%-difference between $\hat{q}(10)$ and $\hat{q}(90)$. This heterogeneity is mostly visible in the right tail—the skewness of $\hat{\phi}^*(a)$ is equal to 6.2 which implies that a minority of funds exhibit stellar investment skills. An open question is whether these skilled funds diversify their holdings across many stocks or instead take concentrated bets. Consistent with the second interpretation, we find that funds with high fd alphas also tend to have high idiosyncratic risk (i.e., the correlation ρ in the bias correction term equals 0.27).

Adjusting for the EIV bias is essential to infer the *true* distribution $\phi(a)$. As discussed in Section II.C, this adjustment requires that we remove the probability mass from the tails, and transfer it onto the center—slightly to the right because ρ is positive (i.e., $\mu_m + \rho\sigma_m\sigma_s > \mu_m$). Both changes are economically significant and explain why all the estimators directly obtained from $\hat{a}_i$ are biased. For one, the unadjusted distribution $\hat{\phi}(a)$ produces an implausibly large difference of 11.4% between the quantiles $\hat{q}(10)$ and $\hat{q}(90)$, and completely fails to capture the asymmetry in skill (the estimted skewness without the EIV correction is a mere 0.5).

Turning to the analysis of the style groups, we document sharp differences between the two size groups (Table III and Figure 4). Specifically, small cap funds produce a

higher average annual fd alpha (5.1% vs 2.0%) and a higher skill proportion $\hat{\pi}_\alpha^+$ (95.4% vs 82.3%). There are two potential explanations for this skill gap. First, small cap stocks have higher idiosyncratic volatility and thus a larger flow of company specific information. Therefore, they provide more opportunities for stock picking activities (e.g., Duan, Yu, and McLean (2009)). Second, their mispricing is more persistent because there is less competitive pressure among funds.[21]

Finally, we measure skill among funds with different levels of expense ratios and turnover. This analysis is motivated by the previous literature which commonly uses both characteristics as predictors of mutual fund performance.[22] Table III shows that high expense funds unambiguously dominate low expense funds, i.e., $\hat{\phi}^*(a)$ exhibits a higher average (4.6% vs 2.2%) and a higher estimated skewness (5.9 vs 5.0). In contrast, trading activity provides little information about the fund investment skills as the proportions $\hat{\pi}_a^-$ and $\hat{\pi}_a^+$ remain largely unchanged. This result suggests that the root cause for the existence of unskilled funds ($\hat{\pi}_a^- = 12.5\%$) is not the overconfidence and excessive trading behaviour of managers. A more plausible explanation is that these funds take active positions to hide their lack of skill from investors (Berk and van Binsbergen (2018)).

Please insert Table III here

Please insert Figure 2 here

A.2 The Size Coefficient

We now turn to the analysis of the size coefficient which measures the fund sensitivity to capacity constraints. To ease interpretation, we compute the standardized size coefficient $\hat{b}_i$ of each fund in Equation (1) so that it corresponds to the annual change in gross alpha for a one standard deviation change in fund size. Because we estimate the entire cross-sectional distribution $\phi(b)$, we contribute to the previous literature that imposes a common size coefficient across funds ($b_i = \bar{b}$).

Table IV shows that 85.9% of the funds in the population are exposed to capacity constraints. The magnitude of the size coefficient is typically large—on average, a one standard deviation increase in fund size reduces the gross alpha by 1.5% per year.[23]

---

[21] Small cap stocks are prone to idiosyncratic fluctuations and asymmetric information problems and are thus largely untouched by mutual funds (e.g., Grompers and Metrick (2001), Hong, Lim, and Stein (2000)).

[22] An non-exhaustive list include Elton et al. (1993)), Gil-Bazo and Ruiz-Verdu (2009), Grinblatt and Titman (1989), Pastor, Stambaugh, and Taylor (2015), and Wermers (2000).

[23] This number is similar to the estimate reported in the recent study by Harvey and Liu (2018b).

These results provide strong support to previous theoretical work that emphasizes the importance of capacity constraints for the fund industry (e.g., BG, Pastor and Stambaugh (2012)).

The distribution $\hat{\phi}^*(b)$ exhibit similar features as $\hat{\phi}^*(a)$. First, it implies a strong heterogeneity in the population, as captured by the 3.6% difference between the quantiles $\hat{q}(90)$ and $\hat{q}(10)$. Therefore, assuming a constant $\bar{b}$ provides fails to capture the diverse impact of capacity constraints across funds. Second, $\hat{\phi}^*(b)$ is heavily skewed with a level of 8.1. This result constrasts with the parametric approach of Harvey and Liu (2018b) in which $\phi(b)$ is assumed to be normal. Intuitively, a normal distribution fails to capture the asymmetric nature of capacity constraints—whereas b_i is bounded around zero for unconstrained funds, it rises significantly for funds constrained by capacity constraints. Finally, the EIV bias adjustment has a large impact—in particular, it allows us to uncover the strong asymmetry of the *true* distribution $\phi(b)$.

The impact of capacity constraints varies significantly across the two size groups (Table IV and Figure 4). For one, small-cap funds have, on average, a higher size coefficient (2.0% vs 1.0%). This finding resonates with previous studies that establish a link between the tightness of capacity constraints and the illiquidity of small-cap stocks (e.g., Chen et al. (2004), Yan (2004)).[24] We also document a similar pattern for groups sorted on characteristics—funds with high expense ratios and high turnover tend to follow strategies that are difficult to scale up. Part of this difficulty is due to the fact that these funds tilt their portfolios toward small cap stocks (see Table II).

Please insert Table IV here

Please insert Figure 3 here

A.3 The Correlation between The two Skill Dimensions

An important insight from our joint analysis of the two skill dimensions is that they are positively correlated. A clear example is provided by the two size groups. We find that small cap funds exhibit both higher fd alphas and higher size coefficients than large cap

---

They find that a \$100 mio. increase in size reduces the gross alpha by 0.17% per year. In our case, the implied reduction is equal to $\frac{\bar{b}}{\bar{\sigma}_q}100$ mio$= \frac{1.5}{6.5} = 0.23\%$ per year, where $\bar{b}, \bar{\sigma}_q$ are the average size coefficient and standard deviation of fund size equal to 1.5% and \$650 mio., respectively.

[24] To check that our results are not driven by the omitted exposure of small cap funds to aggregate liquidity risk, we use an alternative benchmark model that includes the traded liquidity factor of Pastor and Stambaugh (2003). Our results reported in the appendix remain unchanged.

funds. We observe the same empirical regularity among high and low expense funds.[25]

The implications of this positive correlation are twofold. First and foremost, we need to combine the two skill dimensions to determine the overall skill level. For instance, large cap funds may well dominate small cap funds if they are more able to scale up their less profitable trading ideas. We address this issue below by computing the value added which allows us aggregate the two skill dimensions into a single measure.

Second, it helps us reconcile the literature on the predictive content of expense ratios. On the one hand, some studies argue that high expense ratios signal superior skill (e.g., Kacperczyk, van Nieuwerburgh, and Veldkamp (2014), Pastor, Stambaugh, and Taylor (2017)). On the other hand, previous work provides empirical evidence that favors the opposite interpretation (e.g., Elton et al. (1993), Gil-Bazo and Ruiz-Verdu (2009)). Our results show that both conclusions hold depending on which skill dimension is examined—high expense funds are skilled at generating profitable ideas, but unskilled at resisting capacity constraints.

B The Value Added

B.1 The Two Measures of the Value Added

We now examine the two measures of the fund value added. The first measure $va_{i,l}$ determines the profits earned the fund during its entire lifecycle (Equation (2)). The second measure $va_{i,ss}$ determines the profits once the fund reaches its average size (Equation (3)). Each estimated measure is a function of the fd alpha $\hat{a}_i$ and size coefficient $\hat{b}_i$ and thus provides an economically-motivated approach for aggregating the two skill dimensions. The summary statistics for the bias-adjusted densities of the two measures $\hat{\phi}^*(va_l)$ and $\hat{\phi}^*(va_{ss})$ are respectively reported in Panels A and B of Table V.

Overall, mutual funds earn large profits through their investment and trading decisions. The lifecycle value added $va_{i,l}$ is, on average, equal to 1.9 mio. per year, and 64.1% of the funds earn positive profits during their lifecycle.[26] Here again, adjusting for the EIV bias is key for measuring skill—the unadjusted estimator of $\pi_{va,l}^+$ is below 50% and thus largely underestimates the ability of funds to create value. The value creation is even more striking with the steady state value added $va_{i,ss}$. Once funds reach their average size, more than 70% of them earn positive profits, and the average value added jumps to 7.1 mio. per year. This 5.2 mio. difference (7.1-1.9) is due to the negative

---

[25] In the entire population, a simple (biased) correlation calculation based on the estimated coefficients $\hat{a}_i$ and $\hat{b}_i$ yields a large coefficient of 0.84.

[26] The average value added is qualitatively similar to the estimate of 3.2 mio. per year reported by Berk and van Binsbergen (2015) for a sample of US and international funds.

correlation between the gross alpha and size (i.e., $\text{cov}(\alpha_{it}, q_{i,t-1}) < 0$). As funds go through different phases of their lifecycle, their annual profits can be much lower than $va_{i,ss}$ because high values of α_{it} come with low values of $q_{i,t-1}$ (and vice-versa).

We find that the proportion of funds with positive value added is lower than the one obtained with the fd alpha ($\hat{\pi}_{va}^+ - \hat{\pi}_a^+ = 23.4\%$ and $\hat{\pi}_{va_{ss}}^+ - \hat{\pi}_a^+ = 14.8\%$). This decrease arises because some funds grow too large to maintain positive gross alphas. Using this information, we can therefore sort funds with negative value added in two groups: around 40% of them have poor investment ideas, while the remaining 60% are initially skilled but become too large.[27]

Similar to the two skill dimensions, the average value added provides a poor summary of the profits earned across funds. First, the heterogeneity in the value added is substantial—the volatility of $\hat{\phi}^*(va_l)$ and $\hat{\phi}^*(va_{ss})$ is up to 4.8 times larger than its average. Second, both distributions exhibit high levels of skewness and kurtosis. As a result, funds in the right tail reap staggering profits, i.e., $\hat{q}(90)$ reaches 11.1 mio. for $va_{i,l}$ and 25.6 mio. for $va_{i,ss}$. In other words, these funds are particularly skilled and are correctly identified as such by investors.

Turning to the analysis of the different fund groups, we find that small cap funds dominate large cap funds. For both measures of value added, the small cap group produces a higher average (6.3 mio. vs 2.4 mio.) and a higher proportion $\hat{\pi}_{va}^+$ (73.7% vs 60.3%). Therefore, the high fd alphas of small cap funds allow them to reap higher profits even though they face tighter capacity constraints. A similar pattern emerges for low and high turnover funds. Specifically, the low turnover group produces a higher average (9.4 mio. vs 3.5 mio.) and a higher proportion of skilled funds (85.6% vs 64.3%).

The comparison of groups sorted on expense ratios is more subtle. Overall, low expense funds dominate high expense funds—both the average and $\hat{q}(90)$ are significantly larger (7.9 mio. vs 3.4 mio. and 37.5 mio vs 13.4 mio.). However, a higher proportion of low expense funds destroy value once they reach the steady state size (32.5% vs 23.0%). This negative outcome implies that investors allocate too much money to these funds, which then fail to eliminate the negative impact of this excessive inflow.

Please insert Table V here

---

[27] Specifically, we have $\frac{1}{2}\left(\frac{\hat{\pi}_a^-}{\hat{\pi}_{va}^-} + \frac{\hat{\pi}_a^-}{\hat{\pi}_{va_{ss}}^-} \right) = 40.3\%$ and $\frac{1}{2}\left(\frac{\hat{\pi}_{va}^- - \hat{\pi}_a^-}{\hat{\pi}_{va}^-} + \frac{\hat{\pi}_{va_{ss}}^- - \hat{\pi}_a^-}{\hat{\pi}_{va_{ss}}^-} \right) = 59.7\%$.

B.2 Is the Value Added Optimized?

A key prediction of the neoclassical BG model is that skilled funds are in scarce supply and thus able to maximize their profits. In practice, the value added is likely to differ from its optimal value because of learning effects. As investors do not observe the fund skill dimensions a_i and b_i, they must learn about them using past data (Pastor and Stambaugh (2012)). During this learning phase, their fund investments may be very far from the optimal size that maximizes the value added. Whereas the impact of these learning effects should decrease as the fund size approach its average level, they may have a large impact on the average profits earned by funds during their lifecycle.

To address this issue, we use our non-parametric approach to estimate the distribution of the optimal value added va_i^*. Using the estimated coefficients $\hat{a}_i$ and $\hat{b}_i$ from Equation (1), we compute the estimated optimal value added for each fund i as

$$\hat{m}_i = \widehat{va}_i^* = \frac{\hat{a}_i^2}{4\hat{b}_i}, \tag{18}$$

and the error term $u_{i,t}$ as

$$u_{i,t} = \frac{2a_i}{4b_i} e_1' Q_{x,i}^{-1} x_{i,t}\varepsilon_{i,t} - \frac{a_i^2}{4b_i^2} e_2' Q_{x,i}^{-1} x_{i,t}\varepsilon_{i,t}. \tag{19}$$

Then, we apply our methodology to compute the bias-adjusted density $\hat{\phi}^*(va^*)$ and its summary statistics. Equation (20) requires both $\hat{a}_i$ and $\hat{b}_i$ to be positive. Therefore, we conduct the analysis on funds for which this condition is satisfied (75% of the sample).

The empirical evidence in Table VI is broadly supportive of the predictions of the model. The average values for the steady state and optimal value added $va_{i,ss}$ and va_i^* are equal to 9.0 mio. and 13.4 mio. per year. Therefore, funds extract close to 70% of the optimal profits once they reach the steady state size.[28] In contrast, the difference between the lifecyle and optimal value added $va_{i,l}$ and va_i^* is much larger (1.3 mio. vs 13.4 mio.). This result suggests that learning effects have a sizable impact on the profits that funds are able to generate over time.

Consistent with our previous results, we find that low expense and low turnover funds produce the highest optimal value added (23.5 and 27.3 mio., on average). In other words, both groups achieve the most profitable combinations of the two skill dimensions. The biggest gap (in relative terms) between the steady state and optimal

---

[28] In addition, the pairwise correlation between $\widehat{va}_i^*$ and $\widehat{va}_{i,ss}$ equals 0.94, which suggests that funds with the highest skill potential do create more value.

value added is observed among large cap and high turnover funds—in both groups, the ratio $\frac{\widehat{va}_{i,ss}}{\widehat{va}_i^*}$ is barely above 50%, on average. As a result, these funds produce low value added partly because they fail to exploit their skill potential.

<center>Please insert Table VI here</center>

C   The Gross Alpha

C.1   The Information Content of the Gross Alpha

Our final analysis focuses on the gross alpha which is commonly used as a measure of skill. The model presented in Section II.A shows that the gross alpha is related skill only if funds choose specific compensation schemes. Each of them implies specific predictions regarding the levels of fund fees and size (Table I). We can therefore examine whether these predictions hold empirically.

The results in Table VII suggests that the gross alpha contains limited information about skill. First, we find little evidence that funds systematically choose one of the three compensation schemes under which the gross alpha is aligned with the fd alpha, the size coefficient, or the value added. Schemes II (size coefficient) and III (value added) are cleary rejected—the fees are not tiny and the size is not constant across funds. Only Scheme I (fd alpha) is partly supported by the data because both fees and size vary across funds. Second, the pattern for fees and size is consistent with Scheme IV under which the alpha is uninformative about skill. We find that fees and size are negatively correlated (i.e., the pairwise correlation equals -0.19 across funds and -0.96 across fund groups). In other words, some funds are willing to charge low fees in order to manage a large asset base and, possibly, mitigate several institutional constraints (see Habib and Johnson (2016)).[29]

Consistent with this analysis, Table VIII reveals that the summary statistics for the gross alpha distribution $\hat{\phi}^*(\alpha)$ are uninformative about the size coefficient and the value added, and are only weakly related to those reported for the fd alpha (Table III).

<center>Please insert Table VII here</center>

<center>Please insert Table VIII here</center>

---

[29] The Investment Company Act imposes diversification rules on 75% of the portfolio that prevent managers from exhausting their trading opportunities if the fund is too small. Holding a portion of the fund passively managed also allows funds to hide their informed trades and obtain better prices.

<center>25</center>

C.2 From Gross to Net Alphas

Finally, we apply the non-parametric approach to the net alpha. This performance analysis reported in Table IX determines whether investors receive any surplus alphas and provides information about the bargaining power of funds when setting fees.

The neoclassical BG model predicts that the net alphas equal zero as skilled funds are able to extract all the profits they generate (i.e., $\alpha_i = f_{e,i}$). Consistent with this prediction, the proportion of funds with positive alphas drops from 72.6% to 35.4% as move from gross to net alphas.[30,31] In addition, fund groups with higher gross alphas also tend to charge higher fees (the pairwise correlation equals 0.57).

However, we find that the left tail of the net alpha distribution does not shrink towards zero—$\hat{\phi}^*(\alpha^n)$ is essentially a shifted version of $\phi(\alpha)$ with a similar volatility and a fatter left tail. In other words, some funds are able to charge excessively high fees to investors (Christoffersen and Musto (2002), Gruber (1996)). The behaviour of these apparently irrational investors drives a wedge between gross alphas and fees that is left unexplained by the neoclassical model. The same behavior also explains why some unskilled funds are allowed to grow sufficiently large and destroy a sizeable amount of value (Table V).

Please insert Table IX here

D Sensitivity Analysis

D.1 Alternative Asset Pricing Models

(Carhart model). We now summarize the additional results reported in the appendix. First, we use the Carhart model to estimate fund skill. Overall, our analysis reveals that the distributions of the two skill dimensions remain largely unchanged. The fd alpha is equal to 2.7% per year, on average, and is positive for 83.0% of the funds (vs 3.5% and 87.5% for the CPZ model). For the size coefficient, the similarity is even more striking—on average, it is equal to 1.4% per year and 83.9% of the funds have a positive coefficient (vs 1.5% and 85.9%). The main difference is observed for the small cap group

---

[30] Both proportions are significantly larger than those obtained with the False Discovery Rate approach (Barras, Scaillet, and Wermers (2010)). Intuitively, this approach estimates π_α^+ and $\pi_{\alpha^n}^+$ by counting funds with large alpha t-statistics and thus does not detect all the funds with positive gross/net alphas (see Barras (2018)). In contrast, our non-parametric approach directly estimates the gross alpha distribution using information from the entire cross-section of funds.

[31] The presence of search costs for which investors need to be compensated can rationalize the existence of positive net alphas (Garleanu and Pedersen (2018)).

in which the annual fd alpha drops 5.0% to 3.2%, on average. This sharp reduction arises because the Carhart model assigns a negative alpha to the Russell 2000 index.

(Five-factor model). We also measure fund skill using the five-factor model of Fama and French (2015) which includes the market, size, value, profitability, and investment factors. Using this model leaves the distribution of the size coefficient largely unchanged. However, it lowers the proportion of funds with a positive fd alpha from 87.5% to 75.6%. In other words, around 12% of the funds achieve a positive fd alpha because they implement profitability- and investment-based strategies. This tilt is particulary pronounced for low expense and low turnover funds as their average fd alphas respectively drop by 0.9% and 1.3% per year.

(CPZ-liquidity model). Finally, we check that the strong differences between small cap and large cap funds along the two skill dimensions are not simply due to an omitted liquidity risk factor. To address this issue, we add the traded liquidity factor of Pastor and Stambaugh (2003) to the CPZ model. We find that the distributions of the two skill dimensions remain largely unchanged. Therefore, the superior fd alphas produced by small cap funds is not driven by an aggregate liquidity risk premium.

D.2 Small Sample Bias

We also use an alternative estimation procedure to control for the small sample bias in the estimated size coefficient. This bias stems from the positive correlation between the return residual $\varepsilon_{i,t}$ and the change in size $\varepsilon_{qi,t}$, as positive return surprises increase the fund size (e.g., Pastor, Stambaugh, and Taylor (2015), Stambaugh (1999)). Whereas this bias vanishes asympotically, it can potentially affect the estimated coefficients of funds with short return histories. To address this issue, we use the approach of Amihud and Hurvich (2004) that adds a proxy for the size innovation $\hat{\varepsilon}_{qi,t}$ to the set of regressors in Equation (1) (see the appendix for details).[32] The empirical results reveal the bias adjusted densities for the two skill dimensions remain largely unchanged.

D.3 Alternative Specifications for Size

(Industry-adjusted size). We consider different specifications to model the relation between the fund gross alpha $\alpha_{i,t}$ and its size $q_{i,t-1}$. Following Harvey and Liu (2018b), we replace $q_{i,t-1}$ with $q_{i,t-1}^{rel}$, defined as the ratio between the size of the fund and that of the active fund industry. This choice is motivated by the possibility that the fund capacity

---

[32] The inclusion of the estimated variable $\hat{\varepsilon}_{qi,t}$ does not change the properties of our non-parametric density estimator. In particular, the smoothing and EIV biases in Equations (9)-(10) remain unchanged because $\hat{\varepsilon}_{qi,t}$ only affects the higher order terms beyond T^{-1}.

constraints are affected by the total size of the industry. We find that the bias-adjusted distribution $\hat{\phi}^*(b)$ remains largely unchanged.

(*Log size*). Next, we replace $q_{i,t-1}$ with $\log(q_{i,t-1})$ and find qualitatively similar results to those reported in Table IV for the size coefficient. Using a log specification is useful when imposing a constant size coefficient, as fund alphas are more likely to respond similarly to relative changes in size (vs absolute changes). However, this specification seems less important here because we allow for fund-specific size coefficients.

(*Fixed costs*). We also test whether the relation between $\alpha_{i,t}$ and $q_{i,t-1}$ is subject to non-linearities that may bias the estimated size coefficient. One main source of non-linearities is the existence of fixed operating costs F_i. In this case, we have: $\alpha_{i,t} = a_i - b_i q_{i,t-1} - \frac{F_i}{q_{i,t-1}} = a_i - b_i q_{i,t-1} - c_i \frac{1}{q_{i,t-1}}$. Estimating this equation for each fund, we find that the distribution $\hat{\phi}^*(b)$ remains largely unchanged.

(*Industry-wide capacity constraints*). Finally, we re-estimate Equation (1) after replacing $q_{i,t-1}$ with an industry wide measure of capacity constraints, which is computed as the ratio of the industry size on the total market capitalization. Consistent with Pastor, Stambaugh, and Taylor (2015), we find evidence that the individual fund gross alphas vary with industry wide capacity constraints. However, the estimated coefficients are less precisely estimated because the regression condition numbers in Equation (5) increase significantly (the sample size drops to 486 funds).

V   Conclusion

In this paper, we apply a new approach for measuring the entire skill distribution across mutual funds. Our approach is non-parametric and thus particularly suited to the analysis of skill. For one, it avoids the challenge of correctly specifying the skill distribution. It also allows us to jointly examine multiple skill measures, including the two skill dimensions (the fd alpha and size coefficient), and the two measures of the value added (lifecyle and steady state). In addition to its flexibility, our approach brings several advantages—it is simple to implement, applicable to all the descriptive statistics of the skill distribution, and supported by econometric theory.

Our empirical analysis yields several insights into the skill level in the fund industry. It reveals that most funds are able to detect profitable trades—around 85% of them produce positive fd alphas. However, most funds are are also heavily exposed to capacity constraints—around 85% of them have a positive size coefficient. We also find that both skill dimensions are positively correlated. For one, small cap funds produce higher fd alphas than large cap funds, but also are also more exposed to capacity constraints.

This positive correlation provides a strong motivation of examining the value added because this measure combines the two skill dimensions together. Overall, we find that funds earn substantial profits from exploiting their skills—around 70% of them generate positive profits. Finally, the skill distributions for all measures are highly-non normal and reveal substantial heterogeneity across funds. Therefore, computing simple skill averages provides a poor summary of the diversity encountered at the individual fund level.

Whereas our paper focuses on skill, our non-parametric approach has potentially wide applications in finance and economics. We can use it to estimate the cross-sectional distribution of any coefficient of interest in a random coefficient model. This is, for instance, the case in asset pricing when we want to capture the heterogeneity across stocks (e.g., risk exposure, commonality in liquidity), or in corporate finance when we want to capture the heterogeneity across firms (e.g., investment and financing decisions).

References

[1] Ait-Sahalia, Y., 1996, Nonparametric Pricing of Interest Rate Derivative Securities, Econometrica, 1996, 527-560.

[2] Ait-Sahalia, Y., A. Lo, 1998, Nonparametric Estimation of State-Price Densities Implicit in Financial Asset Prices, Journal of Finance, 53, 499-547.

[3] Amihud Y., and C. M. Hurvich, 2004, Predictive Regressions: A Reduced-Bias Estimation Method, Journal of Financial and Quantitative Analysis, 813-841.

[4] Avramov D., L. Barras, and R. Kosowski, 2013, Hedge Fund Return Predictability under the Magnifying Glass, Journal of Quantitative and Financial Analysis 48, 1057-1083.

[5] Baks, K. P., A.Metrick, and J. Wachter, 2001, Should Investors Avoid all Actively Managed Mutual Funds? A Study in Bayesian Performance Evaluation, Journal of Finance 56, 45-85.

[6] Barras L., 2018, A Large-Scale Approach for Evaluating Asset Pricing Models, Journal of Financial Economics, Forthcoming.

[7] Barras L., O. Scaillet, and R. Wermers, 2010, False Discoveries in Mutual Fund Performance: Measuring Luck in Estimated Alphas, Journal of Finance 65, 179-216.

[8] Belsley, D. A., E. Kuh, and R.E., Welch, 2004, Regression Diagnostics—Identifying Influential Data and Sources of Collinearity. New York: Wiley.

[9] Berk J. B., and J. van Binsbergen, 2015, Measuring Skill in the Mutual Fund Industry, Journal of Financial Economics 118, 1-20.

[10] Berk J. B., and J. van Binsbergen, 2018, Regulation of Charlatans in High-Skill Professions, Working Paper.

[11] Berk, J.B., and R.C. Green, 2004, Mutual Fund Flows and Performance in Rational Markets, Journal of Political Economy 112, 1269-1295.

[12] Bhattacharya P. K., 1967, Estimation of a Probability Density Function and its Derivatives, Sankhya Series A 29, 373-382.

[13] Carhart, M. M., 1997, On Persistence in Mutual Fund Performance, Journal of Finance 52, 57-82.

[14] Chen, J., 2017, On Finite Mixture Models, Statistical Theory and Related Fields 1, 15-27.

[15] Chen, Y., M. Cliff, and H. Zhao. 2017. Hedge Funds: The Good, the Bad, and the Lucky. Journal of Financial and Quantitative Analysis 52, 1081–1109

[16] Chen J., H. Hong, M. Huang, and J. D. Kubik, 2004, Does Fund Size Erode Mutual Fund Performance? The Role of Liquidity and Organization, American Economic Review 94, 1276-1302.

[17] Christoffersen, S.E. K., and D. K. Musto, 2002, Demand Curves and the Pricing of Money Management, Review of Financial Studies 15, 1495-1524.

[18] Cremers, M., A. Petajisto, and E. Zitezewitz, 2012, Should Benchmark Indices Have Alpha? Revisiting Performance Evaluation, Critical Finance Review 2012, 1–48.

[19] Duan, Y., G. Hu, R.D. and McLean, 2009, When is Stock Picking Likely to be Successful? Evidence from Mutual Funds, Financial Analysts Journal 65, 1-11.

[20] Elton, E J., M. J. Gruber, S. Das, and M. Hlavka, 1993, Efficiency with Costly Information: A Reinterpretation of Evidence from Managed Portfolios, Review of Financial Studies 6, 1-22.

[21] Elton, E.J., M. J. Gruber, and J. Busse, 2004, Are Investors Rational? Choices among Index Funds, Journal of Finance 59, 261-288.

[22] Ferson, W., and Y. Chen. 2015. How Many Good and Bad Fund Managers are There, Really? Working Paper.

[23] Gagliardini P., E. Ossola, and O. Scaillet, 2016, Time-varying Risk Premium in Large Cross-Sectional Equity Datasets, Econometrica 84, 985-1056.

[24] Gagliardini P., E. Ossola, and O. Scaillet, 2018, A Diagnostic Criterion for Approximate Factor Structure, Working Paper.

[25] Garleanu, N., and L. Pedersen, 2018, Efficiently Inefficient Markets for Assets and Asset Management, Journal of Finance 73, 1663-1712.

[26] Gil-Bazo, J., and P. Ruiz-Verdu, 2009, The Relation between Price and Performance in the Mutual Fund Industry, Journal of Finance 53, 2153-2183.

[27] Ghosh J., and P. Sen, 1985, On the Asymptotic Performance of the Log Likelihood Ratio Statistic for the Mixture Model and Related Results, in Cam, L. L. and Olshen, R., editors, Proc. Berkeley Conf. in honor of Jerzy Neyman and Jack Kiefer, 789-806.

[28] Gompers, P. A., and A. Metrick, 2001, Institutional Investors and Equity Prices, Quarterly Journal of Economics, 116, 229–259.

[29] Greene, W., 2008, Econometric Analysis. Upper Saddle River: Prentice Hall.

[30] Grinblatt, M., and S. Titman, 1989, Mutual Fund Performance: An Analysis of Quarterly Portfolio Holdings, Journal of Business 62, 394–415.

[31] Gruber, M.J., 1996, Another Puzzle: The Growth in Actively Managed Mutual Funds, Journal of Finance 51, 783-810.

[32] Habib, M., and D.B. Johnson, 2016, The Quality-Assuring Role of Mutual Fund Advisory Fees, International Review of Law and Economics, 46, 1-19.

[33] Harvey, C., and Y. Liu, 2018, Detecting Repeatable Performance, Review of Financial Studies, Forthcoming.

[34] Harvey, C., and Y. Liu, 2018b, Decreasing Returns to Scale, Fund Flows, and Performance, Working Paper.

[35] Hong, H., T. Lim, and J. Stein, 2000, Bad News Travels Slowly: Size, Analyst Coverage, and the Profitability of Momentum Strategies, Journal of Finance 55, 265–295.

[36] Hsiao C., 2003, Analysis of Panel Data, Cambridge University Press.

[37] Jensen, M. C., 1968, The Performance of Mutual Funds in the Period 1945-1964, Journal of Finance 23, 389-416.

[38] Jones, C.S., and J. Shanken, 2005, Mutual Fund Performance with Learning across Funds, Journal of Financial Economics 78, 507-552.

[39] Newey, W.K., and K. D. West, 1987, A Simple, Positive Semi-Definite, Heteroscedasticity and Autocorrelation Consistent Covariance Matrix, Econometrica 55, 703-708.

[40] Kacperczyk, M., S. van Nieuwerburgh, and L. Veldkamp, 2014, Time-Varying Fund Manager Skill, Journal of Finance 69, 1455-1484.

[41] Okui, R., and T. Yanagi, 2018, Kernel Estimation for Panel Data with Heterogeneous Dynamics, Working Paper.

[42] Pastor L., and R. F. Stambaugh, 2003, Liquidity Risk and Expected Stock Returns, Journal of Political Economy 111, 642-685.

[43] Pastor, L., and R. F. Stambaugh, 2012, On the Size of the Active Management Industry, Journal of Political Economy 120, 740-781.

[44] Pastor, L., R. F. Stambaugh, and L. Taylor, 2015, Scale and Skill in Active Management, Journal of Financial Economics 116, 23-45.

[45] Pastor, L., R. F. Stambaugh, and L. Taylor, 2016, Do Funds Make More When They Trade More?, Journal of Finance, Forthcoming.

[46] Perold, A. F., and R. S. Salomon, Jr., 1991, The Right Amount of Assets under Management, Financial Analysts Journal 47, 31-39.

[47] Pollet, J.,and M. Wilson,2008, How does Size Affect Mutual Fund Behavior? Journal ofFinance 63, 2941–2969.

[48] Shanken J., 1992, On the Estimation of Beta-Pricing Models, Review of Financial Studies 5, 1-33.

[49] Silverman, B.W., 1986, Density Estination for Statistics and Data Analysis, Chapman and Hall, London.

[50] Stambaugh,R.F.,1999, Predictive Regressions, Journal of Financial Economics 54, 375–421.

[51] Stambaugh, R.F, Presidential Address: Investment Noise and Trends, Journal of Finance 69, 1415-1453.

[52] Stanton, R., 1997, A Nonparametric Model of Term Structure Dynamics and the Market Price of Interest Rate Risk, Journal of Finance 52, 1973-2002.

[53] Wand, M.P., and M.C., Jones, 1995, Kernel Smoothing, Chapman and Hall, London.

[54] Wermers, Russ, 2000, Mutual Fund Performance: An Empirical Decomposition into Stock-Picking Talent, Style, Transaction Costs, and Expenses, Journal of Finance 55, 1655-1695.

[55] Yan, X., 2008, Liquidity, Investment Style, and the Relation between Fund Size and Fund Performance, Journal of Financial and Quantitative Analysis 43,741–768.

VI Appendix

A.1 Estimation of the density of the skill measures

In this appendix, we focus on the proof of Proposition II.1 stated for the four skill measures. The proof for smoothing estimated slopes (betas) or other quantities estimated with parametric rates follows similar arguments. Let us first focus on the fd alpha $m_i = a_i$. From the OLS estimation, we have:

$$
\begin{aligned}
\hat{m}_i &= e_1' \hat{Q}_{x,i}^{-1} \frac{1}{T_i} \sum_t I_{i,t} x_t r_{i,t} = m_i + e_1' \hat{Q}_{x,i}^{-1} \frac{1}{T_i} \sum_t I_{i,t} x_t \varepsilon_{i,t} \\
&= m_i + \frac{1}{\sqrt{T}} \tau_{i,T} e_1' \hat{Q}_{x,i}^{-1} \left(\frac{1}{\sqrt{T}} \sum_t I_{i,t} x_t \varepsilon_{i,t} \right) \equiv m_i + \frac{1}{\sqrt{T}} \hat{\eta}_{i,T}.
\end{aligned} \tag{20}
$$

Moreover, let us write

$$\hat{\eta}_{i,T} = \eta_{i,T} + \frac{1}{\sqrt{T}}\hat{v}_{i,T}, \tag{21}$$

where $\eta_{i,T} = \tau_i \frac{1}{\sqrt{T}}\sum_t I_{i,t}u_{i,t}$, $u_{i,t} = e_1' Q_x^{-1} x_t \varepsilon_{i,t}$, and $\tau_i = \text{plim}_{T\to\infty}\tau_{i,T}$. Hence, $\hat{\eta}_{i,T}/\sqrt{T}$ is the estimation error on $m_i = \alpha_i$. In particular, $\hat{v}_{i,T}/T$ is the component due to estimating the matrix Q_x and to the random sample size T_i. Then, we write $\hat{\phi}(m) - \phi(m) = I_1 + I_2 + I_3 + I_4$, where:

$$I_1 = \frac{1}{h}E\left[K\left(\frac{m_i - m}{h}\right)\right] - \phi(m),$$

$$I_2 = \frac{1}{h}E\left[K\left(\frac{m_i + \eta_{i,T}/\sqrt{T} - m}{h}\right)\right] - \frac{1}{h}E\left[K\left(\frac{m_i - m}{h}\right)\right],$$

$$I_3 = \frac{1}{nh}\sum_i\left\{K\left(\frac{m_i + \eta_{i,T}/\sqrt{T} - m}{h}\right) - E\left[K\left(\frac{m_i + \eta_{i,T}/\sqrt{T} - m}{h}\right)\right]\right\},$$

$$I_4 = \frac{1}{nh}\sum_i\left[\mathbf{1}_i^\chi K\left(\frac{m_i + \eta_{i,T}/\sqrt{T} + \hat{v}_{i,T}/T - m}{h}\right) - K\left(\frac{m_i + \eta_{i,T}/\sqrt{T} - m}{h}\right)\right].$$

The first term I_1 is the smoothing bias, the second term I_2 is the Error-in-Variable (EIV) bias, I_3 is the main stochastic term, and I_4 is a remainder term, which is negligible w.r.t. the others. We characterise the first three dominating terms in the following.

(i) From standard results in kernel density estimation, the smoothing bias is such that $I_1 = \frac{1}{2}\phi^{(2)}(m)K_2h^2 + O(h^3)$, with $K_2 = \int u^2 K(u)du$.

(ii) By a Taylor expansion of the kernel function K we have:

$$I_2 = \sum_{j=1}^\infty \frac{1}{j!T^{j/2}h^{j+1}}E\left[K^{(j)}\left(\frac{m_i - m}{h}\right)\eta_{i,T}^j\right].$$

Moreover, by applying j times partial integration and a change of variable:

$$\begin{aligned}
\frac{1}{h^{j+1}}E\left[K^{(j)}\left(\frac{m_i - m}{h}\right)\eta_{i,T}^j\right] &= \frac{1}{h^{j+1}}\int K^{(j)}\left(\frac{u - m}{h}\right)\psi_{T,j}(u)du \\
&= (-1)^j\frac{1}{h}\int K\left(\frac{u - m}{h}\right)\psi_{T,j}^{(j)}(u)du \\
&= (-1)^j\int K(u)\psi_{T,j}^{(j)}(m + hu)du,
\end{aligned}$$

where $\psi_{T,j}(m) = E[\eta_{i,T}^j | m_i = m]\phi(m)$ for $j = 1, 2, ...$ We have $\psi_{T,1}(m) = 0$ and

34

$\lim_{T\to\infty}\psi_{T,2}(m) = E[S_i|m_i = m]\phi(m) \equiv \psi(m)$ where $S_i = \tau_i^2\mathrm{plim}_{T\to\infty}\frac{1}{T}\sum_{t,s}I_{i,t}I_{i,s}u_{i,t}u_{i,s}$.

By weak serial dependence of the error terms, functions $\psi_{T,j}(m)$ for $j > 2$ are bounded with respect to T. Thus, we get: $I_2 = \frac{1}{2T}\psi^{(2)}(m) + O(1/T^{3/2} + h^2/T)$.

(iii) Let us now consider term I_3. For expository purpose, let us assume that the error terms are cross-sectionally independent, and the factor values x_t are treated as given constants. Then:

$$V[I_3] = \frac{1}{nh^2}V\left[K\left(\frac{m_i + \eta_{i,T}/\sqrt{T} - m}{h}\right)\right].$$

From the above arguments, we have $\frac{1}{h}E\left[K\left(\frac{m_i + \eta_{i,T}/\sqrt{T} - m}{h}\right)\right] = \phi(m) + o(1)$ and

$$\frac{1}{h}E\left[K\left(\frac{m_i + \eta_{i,T}/\sqrt{T} - m}{h}\right)^2\right] = \int K(u)^2 du \frac{1}{h}E\left[\bar{K}\left(\frac{m_i + \eta_{i,T}/\sqrt{T} - m}{h}\right)\right]$$

$$= \phi(m)\int K(u)^2 du + o(1),$$

where $\bar{K}(u) = K(u)^2/\int K(u)^2 du$. Therefore:

$$V[I_3] = \frac{1}{nh}\phi(m)\int K(u)^2 du + o\left(\frac{1}{nh}\right).$$

Under regularity conditions, the application of an appropriate Central Limit Theorem (CLT) implies $\sqrt{nh}I_3 \Rightarrow N\left(0, \phi(m)K_1\right)$, with $K_1 = \int K(u)^2 du$.

For the a (fd dollar), b (size coefficient), and va (value added) measures, we can proceed similarly by using the corresponding definition for $u_{i,t}$ listed in the statement of Proposition I.1.

A.2 Asymptotic Mean Integrated Squared Error

From the previous subsection, we get the asymptotic expansion of the bias and variance of the estimator $\hat{\phi}(m)$ with leading terms: $bs(m) = bs_1(m) + bs_1(m)$, and $\sigma^2(m) =$

$\frac{1}{nh}\phi(m)K_1$. The AMISE (asymptotic mean integrated squared error) is given by:

$$
\begin{aligned}
AMISE(h) &= \int [\sigma^2(u) + bs(u)^2]du \\
&= \frac{1}{nh}K_1 + \frac{h^4 K_2^2}{4}\int [\phi^{(2)}(u)]^2 du \\
&\quad + \frac{h^2 K_2}{2T}\int \phi^{(2)}(u)\psi^{(2)}(u)du + \frac{1}{4T^2}\int \left[\psi^{(2)}(u)\right]^2 du.
\end{aligned}
$$

The optimal bandwidth h^* is the minimizer of the AMISE, and solves the equation:

$$
-\frac{1}{nh^2} + c_1 h^3 + c_2\frac{h}{T} = 0
$$
$$
\iff \quad 1 = c_1 nh^5 + c_2\frac{nh^3}{T}. \tag{22}
$$

where $c_1 = K_2^2 \int [\phi^{(2)}(u)]^2 du/K_1$ and $c_2 = K_2 \int \phi^{(2)}(u)\psi^{(2)}(u)du/K_1$.

Let us investigate the speed of convergence to 0 of the optimal bandwidth h^* as a function of n and T. We assume that $c_2 > 0$. There are three possible cases:

(i) The optimal bandwidth is such that nh^5 tends to a nonzero constant and $nh^3/T \to 0$. Then, we have $h^* \sim c_1^{-1/5}n^{-1/5}$, that is the Silverman rule. This solution is admissible, i.e., satisfies $nh^3/T \to 0$, if the sample sizes n and T are such that $n^{2/5}/T \to 0$, i.e., T is sufficiently large.

(ii) The optimal bandwidth is such that nh^3/T tends to a nonzero constant and $nh^5 \to 0$. This case is possible only if $n/T \to \infty$. Then, we have $h^* \sim c_2^{-1/3}(n/T)^{-1/3}$. This solution is admissible, i.e., satisfies $nh^5 \to 0$, if the sample sizes are such that $n^{2/5}/T \to \infty$.

(iii) When $n^{2/5}/T \to \rho$, with $\rho > 0$, the two rates of convergence $n^{-1/5}$ and $(n/T)^{-1/3}$ coincide. Then, equation (22) has a solution such that $h^* \sim \bar{c}^{1/5}n^{-1/5}$, where $\bar{c}$ solves the equation $1 = c_1\bar{c} + c_2\rho\bar{c}^{3/5}$.

Let us now consider the asymptotic distribution of estimator $\hat{\phi}(m)$ for a generic bandwidth sequence h shrinking to zero such that $nh \to \infty$. From the above analysis, we have:

$$
\sqrt{nh}\left(\hat{\phi}(m) - \phi(m) - bs(m)\right) \Rightarrow N\left(0, \phi(m)K_1\right).
$$

For some bandwidth sequences, the asymptotic bias is negligible. (i) If $Th^2 \to \infty$, the dominant component in the asymptotic bias is due to smoothing and is of order $O(h^2)$.

Then, the asymptotic bias is negligible if $nh^5 \to 0$. This condition is compatible with the condition $Th^2 \to \infty$ if $n/T^{5/2} \to 0$. (ii) If $Th^2 \to 0$, the dominant component in the asymptotic bias is due to the EIV problem and is of order $O(1/T)$. The asymptotic bias is negligible if $nh/T^2 \to 0$.

Let us now consider the asymptotic distribution when $h = h^*$ is the optimal bandwidth. We can check that $\sqrt{nh^*}(h^{*3} + h^{*2}/T + 1/T^{3/2}) = o(1)$ if $n/T^4 \to 0$. Then, we can replace the bias component $bs(m)$ by its asymptotic approximation to get:

$$\sqrt{nh^*}\left(\hat{\phi}(m) - \phi(m) - \frac{1}{2}\phi^{(2)}(m)K_2h^{*2} - \frac{1}{2T}\psi^{(2)}(m)\right) \Rightarrow N\left(0, \phi(m)K_1\right). \quad (23)$$

If $n^{2/5}/T \to \infty$, we have $Th^{*2} \to 0$, and the smoothing bias is negligible. If $n^{2/5}/T \to 0$, we have $Th^{*2} \to \infty$, and the EIV bias is negligible.

A.3 A simple plug-in method

In this section we develop a simple plug-in method to implement the optimal bandwidth that is solution of equation (22). We focus on a Gaussian kernel $K(u) = \frac{1}{\sqrt{2\pi}}\exp\left(-u^2/2\right)$. Then, the kernel constants are $K_1 = \int K(u)^2 du = \frac{1}{2\sqrt{\pi}}$ and $K_2 = \int u^2 K(u)du = 1$. To compute constants c_1 and c_2, we rely on a reference model which assumes a bivariate Gaussian distribution for m_i and $s_i = \log(S_i)$ with mean parameters μ_m, μ_s, variance parameters σ_m^2, σ_s^2, and correlation parameter ρ. The Gaussian marginal density of m_i implies that our reference model nests the one underlying the derivation of the Silverman rule for kernel smoothing. The constants c_1 and c_2 are given by:

$$c_1 = 2\sqrt{\pi}\int [\phi^{(2)}(u)]^2 du,$$

$$c_2 = 2\sqrt{\pi}\int \phi^{(2)}(u)\psi^{(2)}(u)du = 2\sqrt{\pi}\int \phi^{(4)}(u)\psi(u)du,$$

where we have used twice partial integration in c_2. Let us now compute the two integrals appearing in these formulas.

(i) We have $\phi(m) = \frac{1}{\sigma_m}\varphi\left(\frac{m - \mu_m}{\sigma_m}\right)$, where $\varphi(z) = \frac{1}{\sqrt{2\pi}}\exp\left(-z^2/2\right)$ is the standard Gaussian density. We get: $\phi^{(1)}(m) = -\frac{1}{\sigma_m}\left(\frac{m - \mu_m}{\sigma_m}\right)\phi(m)$, and $\phi^{(2)}(m) =$

$$\frac{1}{\sigma_m^2}\left(\left(\frac{m-\mu_m}{\sigma_m}\right)^2-1\right)\phi(m). \text{ Then:}$$

$$
\begin{aligned}
\int[\phi^{(2)}(u)]^2 du &= \frac{1}{\sigma_m^5}\int(z^2-1)^2\frac{1}{2\pi}\exp(-z^2)dz = \frac{1}{2\sqrt{\pi}\sigma_m^5}\int(v^2/2-1)^2\varphi(v)dv \\
&= \frac{3}{8\sqrt{\pi}\sigma_m^5},
\end{aligned}
$$

with the changes of variables from u to $z=(u-\mu_m)/\sigma_m$, and from z to $v=\sqrt{2}z$.

(ii) We have $E[\exp(s_i)|m_i=m]=\exp\left(\mu_s+\rho\sigma_s\left(\frac{m-\mu_m}{\sigma_m}\right)+\frac{1}{2}\sigma_s^2(1-\rho^2)\right)$, so that $\psi(m)=\exp\left(\mu_s+\rho\sigma_s\left(\frac{m-\mu_m}{\sigma_m}\right)+\frac{1}{2}\sigma_s^2(1-\rho^2)\right)\phi(m)$ and

$$
\begin{aligned}
\psi^{(2)}(m) &= \exp\left(\mu_s+\rho\sigma_s\left(\frac{m-\mu_m}{\sigma_m}\right)+\frac{1}{2}\sigma_s^2(1-\rho^2)\right)\phi(m) \\
&\quad \times\left\{\left(\frac{\sigma_s\rho}{\sigma_m}\right)^2-2\frac{\sigma_s\rho}{\sigma_m^2}\left(\frac{m-\mu_m}{\sigma_m}\right)+\frac{1}{\sigma_m^2}\left[\left(\frac{m-\mu_m}{\sigma_m}\right)^2-1\right]\right\} \\
&= \exp\left(\mu_s+\frac{1}{2}\sigma_s^2\right)\frac{1}{\sigma_m^2}\left(\left(\frac{m-\mu_m-\rho\sigma_s\sigma_m}{\sigma_m}\right)^2-1\right)\frac{1}{\sigma_m}\varphi\left(\frac{m-\mu_m-\rho\sigma_s\sigma_m}{\sigma_m}\right).
\end{aligned}
$$

We can also derive directly the last line by rewriting $\psi(m)=\omega(m)\phi(m)$ in the bivariate Gaussian reference model as a recentered Gaussian density up to a multiplicative constant:

$$\psi(m)=\exp\left(\mu_s+\frac{1}{2}\sigma_s^2\right)\frac{1}{\sigma_m}\varphi\left(\frac{m-\mu_m-\rho\sigma_s\sigma_m}{\sigma_m}\right),$$

and then differentiating twice that expression. Besides, we have:

$$
\begin{aligned}
\int\phi^{(4)}(m)\psi(m)dm &= \frac{\exp\left(\mu_s+\frac{1}{2}\sigma_s^2(1-\rho^2)\right)}{\sigma_m^5}\int\varphi^{(4)}(z)\exp(\rho\sigma_s z)\varphi(z)dz \\
&= \frac{\exp\left(\mu_s+\frac{1}{2}\sigma_s^2(1-\rho^2)\right)}{2\sqrt{\pi}\sigma_m^5}\int(v^4/4-3v^2+3)\exp(\lambda v)\varphi(v)dv,
\end{aligned}
$$

where $\lambda=\rho\sigma_s/\sqrt{2}$, by using the same changes of variables as above and $\varphi^{(4)}(z)=(z^4-6z^2+3)\varphi(z)$. Moreover, we have $\int z^k\exp(\lambda z)\varphi(z)dz=E[Z^k\exp(\lambda Z)]=\frac{\partial^k}{\partial\lambda^k}E[\exp(\lambda Z)]$

with $E[\exp(\lambda Z)] = \exp(\lambda^2/2)$ for a standard Gaussian variable Z. This yields

$$\int (v^4/4 - 3v^2 + 3)\exp(\lambda v)\varphi(v)dv = \left(\frac{1}{4}\frac{\partial^4}{\partial\lambda^4} - 3\frac{\partial^2}{\partial\lambda^2} + 3\right)\exp(\lambda^2/2)$$
$$= \frac{1}{4}(\lambda^4 - 6\lambda^2 + 3)\exp(\lambda^2/2).$$

Thus, we get:

$$\int \phi^{(4)}(m)\psi(m)dm = \frac{3\exp\left(\mu_s + \frac{1}{2}\sigma_s^2(1-\rho^2/2)\right)}{8\sqrt{\pi}\sigma_m^5}(\rho^4\sigma_s^4/12 - \rho^2\sigma_s^2 + 1).$$

The optimal bandwidth h^* is obtained by solving Equation (22) with coefficients c_1 and c_2 given by:

$$c_1 = \frac{3}{4\sigma_m^5}, \qquad c_2 = \frac{3}{4\sigma_m^5}(\rho^4\sigma_s^4/12 - \rho^2\sigma_s^2 + 1)\exp\left(\mu_s + \frac{1}{2}\sigma_s^2(1-\rho^2/2)\right).$$

We have $c_2 \geq 0$ when either $\rho^2\sigma_s^2 \leq 6-2\sqrt{6}$, or $\rho^2\sigma_s^2 \geq 6+2\sqrt{6}$. In our implementation, the parameters σ_m, μ_s, σ_s and ρ are estimated by the sample moments of $\hat{m}_i$ and $\hat{s}_i = \log \hat{S}_i$.

A.4 Estimation of the moments of the skill measures

Let us consider the estimation of the cross-sectional expectation $E[g(m_i)]$, where g is a given smooth function. We investigate the convergence properties of the cross-sectional estimator $\frac{1}{n}\sum_{i=1}^{n} g(\hat{m}_i)\mathbf{1}_i^\chi$ based on the OLS estimates $\hat{m}_i$ of the non-trimmed assets only. Proposition VI.1 states the asymptotic normality under the double asymptotics "large n, small T" in unbalanced panel when we assume the linear factor model (2).

Proposition VI.1 *As $n,T \to \infty$, such that $n = o(T^3)$,*

$$\sqrt{n}\left(\frac{1}{n}\sum_i g(\hat{m}_i)\mathbf{1}_i^\chi - E\left[g(m_i)\right] - \mathcal{B}_T\right) \Rightarrow N\left(0, V\left[g(m_i)\right]\right), \tag{24}$$

with $\mathcal{B}_T = \frac{1}{2T}E[g^{(2)}(m_i)S_i].$

Here we have an asymptotic bias $\mathcal{B}_T$ of order $1/T$ where S_i is the same as in Proposition II.1. It comes from the estimation error in $\hat{m}_i$ (EIV contribution). The asymptotic variance is simply the cross-sectional variance $V[g(m_i)]$. The condition $n = o(T^3)$ is

used to control the remainder term in the Taylor expansion of function g, and the bias term. For the mean of the skill measures, we have $g(m) = m$ and the asymptotic bias is zero from $g^{(2)}(m) = 0$. For that particular case, we do not need the condition $n = o(T^3)$ for Proposition VI.1 to hold. For the second moment of the skill measures, we have $g(m) = m^2$, and the asymptotic bias is not zero from $g^{(2)}(m) = 2$. We need to estimate the asymptotic bias and asymptotic variance to get a feasible asymptotic normality result and to build asymptotic confidence intervals. We can achieve that by replacing the unknown moments by consistent estimators based on empirical averages as done earlier in the paper. For the variance of the skill measures, the asymptotic bias is the same as the one for the second moment, but the asymptotic variance is $V\left[(m_i - E\left[m_i\right])^2\right]$. The asymptotic bias and variance for the skewness and kurtosis can be obtained via the delta method. For example, for the skewness $Sk = E\left[(m_i - E\left[m_i\right])^3\right] / E\left[(m_i - E\left[m_i\right])^2\right]^{3/2}$, its asymptotic bias is

$$\mathcal{B}_T(Sk) = (\nabla_3 Sk)\mathcal{B}_T(E[m_i^3]) + (\nabla_2 Sk)\mathcal{B}_T(E[m_i^2]),$$

with

$$\nabla_3 Sk = V[m_i]^{-3/2},$$
$$\nabla_2 Sk = -3E[m_i]V[m_i]^{-3/2} + \left\{E[m_i^3] - 3E[m_i^2]E[m_i] + 2E[m_i^3]\right\}(\frac{-3}{2})V[m_i]^{-5/2}$$

where we denote the derivative of Sk w.r.t. $E[m_i^j]$ as $\nabla_j Sk$. Similarly, for the kurtosis $Ku = E\left[(m_i - E\left[m_i\right])^4\right] / E\left[(m_i - E\left[m_i\right])^2\right]^2$, its asymptotic bias is

$$\mathcal{B}_T(Ku) = (\nabla_4 Ku)\mathcal{B}_T(E[m_i^4]) + (\nabla_3 Ku)\mathcal{B}_T(E[m_i^3]) + (\nabla_2 Ku)\mathcal{B}_T(E[m_i^2]),$$

with

$$\nabla_4 Ku = V[m_i]^{-2}, \qquad \nabla_3 Ku = -4E[m_i]V[m_i]^{-2},$$
$$\nabla_2 Ku = 6E[m_i]^2 V[m_i]^{-2} + \left\{E[m_i^4] - 4E[m_i^3]E[m_i] + 6E[m_i^2]E[m_i]^2 - 3E[m_i]^4\right\}(-2)V[m_i]^{-3}.$$

Proof of Proposition VI.1 : Equation (20) yields the mean value expansion

$$g(\hat{m}_i) = g(m_i) + g^{(1)}(\bar{m}_i)\frac{1}{\sqrt{T}}\hat{\eta}_{i,T} + g^{(2)}(\bar{m}_i)\frac{1}{2T}\hat{\eta}_{i,T}^2,$$

where $\bar{m}_i$ lies between $\hat{m}_i$ and m_i. Then we get

$$\sqrt{n}\left(\frac{1}{n}\sum_i g(\hat{m}_i)\mathbf{1}_i^\chi - E\left[g(m_i)\right] - \mathcal{B}_T\right)$$

$$= \frac{1}{\sqrt{n}}\sum_i\left(g(m_i) - E\left[g(m_i)\right]\right) - \frac{1}{\sqrt{n}}\sum_i g(m_i)(1 - \mathbf{1}_i^\chi) + \frac{1}{\sqrt{nT}}\sum_i \mathbf{1}_i^\chi g^{(1)}(\bar{m}_i)\hat{\eta}_{i,T}$$

$$+ \frac{1}{2T}\frac{1}{\sqrt{n}}\sum_i\left(\mathbf{1}_i^\chi g^{(2)}(\bar{m}_i)\hat{\eta}_{i,T}^2 - E\left[g^{(2)}(m_i)S_i\right]\right)$$

$$\equiv I_{71} + I_{72} + I_{73} + I_{74}.$$

We have $I_{71} \Rightarrow N(0, V[g(m_i)])$ from the standard CLT. We also have $I_{72} = o_p(1)$. The bound $I_{73} = O_p(1/\sqrt{T}) = o_p(1)$ follows from similar arguments as in Lemma 2 of GOS. Then, the asymptotic distribution (24) follows from the remainder term $I_{74} = O_p(\sqrt{n/T^3} + \sqrt{n}/T^2 + 1/T)$, which gives $I_{74} = o_p(1)$ if $n = o(T^3)$.

A.5 Estimation of the cumulative distribution function and quantiles of the skill measures

We consider the estimation of the cumulative distribution function (cdf) $\Phi(m) = P[m_i \leq m]$ of the skill measures, for any given real argument m, and of the associated quantile function $Q(u) = \Phi^{-1}(u)$, for any given percentile level $u \in (0, 1)$, in the linear factor model (2). Building on the previous section, the estimator of the cdf is the cross-sectional average of the indicator function $g(\hat{m}_i) = \mathbf{1}\{\hat{m}_i \leq m\}$ based on the OLS estimates $\hat{m}_i$ for the non-trimmed assets, namely $\hat{\Phi}(m) = \frac{1}{n}\sum_i \mathbf{1}\{\hat{m}_i \leq m\}\mathbf{1}_i^\chi$. The quantile estimator is the inverse function $\hat{Q}(u) = \hat{\Phi}^{-1}(u)$.

The next proposition extends Proposition VI.1 to the estimation of the cdf and quantile function of the skill measures in the linear factor model (2).

Proposition VI.2 *As $n, T \to \infty$, such that $n = o(T^3)$,*

$$\sqrt{n}\left(\hat{\Phi}(m) - \Phi(m) - \mathcal{B}_T(m)\right) \Rightarrow N\left(0, \Phi(m)(1 - \Phi(m))\right),$$

$$\sqrt{n}\left(\hat{Q}(u) - Q(u) + \frac{\mathcal{B}_T(Q(u))}{\phi(Q(u))}\right) \Rightarrow N\left(0, \frac{u(1 - u)}{\phi(Q(u))^2}\right),$$

with $\mathcal{B}_T(m) = \frac{1}{2T}\psi^{(1)}(m)$, where $\psi(m) = E[S_i|m_i = m]\phi(m)$.

As in the previous section, we can approximate the asymptotic bias through a ref-

erence model and estimate the asymptotic variance to get feasible results.[33] With our bivariate Gaussian reference model, we get:

$$
\begin{aligned}
\psi^{(1)}(m) &= \exp\left(\mu_s + \rho\sigma_s\left(\frac{m-\mu_m}{\sigma_m}\right) + \frac{1}{2}\sigma_s^2(1-\rho^2)\right)\phi(m)\left(\frac{\sigma_s\rho}{\sigma_m} - \frac{m-\mu_m}{\sigma_m^2}\right) \\
&= \exp\left(\mu_s + \frac{1}{2}\sigma_s^2\right)\frac{-1}{\sigma_m}\left(\frac{m-\mu_m - \rho\sigma_s\sigma_m}{\sigma_m}\right)\frac{1}{\sigma_m}\varphi\left(\frac{m-\mu_m - \rho\sigma_s\sigma_m}{\sigma_m}\right).
\end{aligned}
$$

We can proceed similarly for the quantile case.

Proof of Proposition VI.2 : From (20), we have: $E\left[1\{\hat{m}_i \le m\}\right] = P\left[m_i + \frac{1}{\sqrt{T}}\hat{\eta}_{i,T} \le m\right]$. By using the results in Gourieroux, Laurent, and Scaillet (2000), Martin, and Wilde (2001), Gordy (2003), Gagliardini and Gourieroux (2011), we get:

$$
\begin{aligned}
P\left[m_i + \frac{1}{\sqrt{T}}\hat{\eta}_{i,T} \le m\right] &= \Phi(m) - \frac{1}{\sqrt{T}}\phi(m)E[\hat{\eta}_{i,T}|m_i = m] \\
&\quad + \frac{1}{2T}\frac{d}{dm}(\phi(m)E[\hat{\eta}_{i,T}^2|m_i = m]) + o(1/T).
\end{aligned}
$$

From (21), the bias expansion is such that: $E[\hat{\Phi}(m)] - \Phi(m) = \mathcal{B}_T(m) + E\left[1\{\hat{m}_i \le m\}(1 - \mathbf{1}_i^\chi)\right] + o(1/T)$. We deduce the asymptotic normality of the cdf estimator by controlling the different terms and applying a CLT.

We deduce the asymptotic normality of the quantile estimator by using the Bahadur expansion for the quantile estimator at level $u \in (0,1)$: $\hat{Q}(u) - Q(u) = -\frac{1}{\phi(Q(u))}\left(\hat{\Phi}(Q(u)) - u\right)$.

---

[33] The asymptotic bias takes the same form as the one in Jochmans and Weidner (2018) where they consider n parameters of interest directly drawn from a Gaussian distribution whose measurement errors decrease at a parametric rate $\sqrt{T}$. In their setting, they use other arguments based on the behaviour of the probability integral transform for their proofs.

Table I
Relations Between the Skill Measures

This table describes different compensation schemes chosen by the funds when setting their level of fees. Each compensation scheme yields specific predictions regarding the cross-sectional distributions of fund fees and size, and the relation between the gross alpha and the three other skill measures (first-dollar alpha, size coefficient, and value added).

Compensation Scheme	Scheme I (optimal size)	Scheme II (squared optimal size)	Scheme III (same size)	Scheme IV (arbitrary size)
Fees are set such that	Managers choose the optimal size	Managers choose the squared optimal size	Managers choose the same average size	Funds choose an arbitrary size
Predictions for Size/Fees	Moderate cross-fund variation in fees/size	Huge size Tiny fees	Same Size for all funds	Large cross-fund variation in size
Does the Gross Alpha Measure Skill?	YES	YES	YES	NO
	First-Dollar Alpha (1st skill dimension)	Size Coefficient (2nd skill dimension)	Value Added	

Table II
Summary Statistics for the Value-Weighted Portfolio of Funds

Panel A reports the average number of funds and the first four moments of the portfolio gross excess return for all funds in the population, four styles groups (small cap, large cap, growth, value), and four characteristic-sorted groups (low expense, high expense, low turnover, high turnover). Panel B reports the estimated portfolio betas on the market, size, value, and momentum factors, as well as the adjusted R^2 using the Cremers, Petajisto, and Zitzewitz benchmark model. All statistics are computed using monthly data between January 1979 and December 2015.

Panel A: Gross Excess Return

	Average Nb. Funds	Mean (Ann.)	Volatility (Ann.)	Skewness	Kurtosis
All Funds	1103	7.68	15.01	-0.8	5.4
Investment Styles					
Small-cap	236	8.82	19.30	-0.6	4.9
Large-cap	436	7.77	14.77	-0.7	5.3
Growth	511	8.14	16.26	-0.8	5.3
Value	279	7.45	13.81	-0.7	5.5
Fund Characteristics					
Low Expense	373	7.71	14.58	-0.8	5.3
High Expense	289	8.03	16.40	-0.8	5.2
Low Turnover	274	7.49	14.31	-0.8	5.5
High Turnover	260	8.58	16.59	-0.7	5.0

Panel B: Estimated Betas

	Market	Size	Value	Momentum	Adj. R2
All Funds	0.93	0.26	-0.10	0.01	0.98
Investment Styles					
Small-cap	0.98	0.81	-0.29	0.05	0.97
Large-cap	0.94	0.15	-0.06	0.01	0.99
Growth	0.95	0.33	-0.27	0.03	0.97
Value	0.92	0.13	0.19	-0.01	0.98
Fund Characteristics					
Low Expense	0.92	0.21	-0.04	0.00	0.98
High Expense	0.93	0.42	-0.27	0.01	0.97
Low Turnover	0.91	0.20	0.04	-0.02	0.98
High Turnover	0.95	0.37	-0.28	0.08	0.97

Table III
Cross-Sectional Distribution of the First Dollar Alpha

The table contains the summary statistics on the cross-sectional distribution of the first dollar (fd) alpha for all funds, four styles groups (small cap, large cap, growth, value), and four characteristic-sorted groups (low expense, high expense, low turnover, high turnover). It reports the first four moments, the proportions of funds with a negative and positive fd alpha, and the distribution quantiles at 10% and 90%. All cross-sectional estimates are adjusted for bias (smoothing and EIV) using our non-parametric approach.

	Moments				Proportions (%)		Quantiles (Ann.)	
	Mean (Ann.)	Volatility (Ann.)	Skewness	Kurtosis	Negative Skill	Positive Skill	10%	90%
All Funds	3.5	3.6	6.2	5.8	12.5	87.5	-0.4	7.8
Investment Styles								
Small-cap	5.1	3.4	10.5	1.5	4.6	95.4	1.2	9.7
Large-cap	2.0	2.4	5.9	3.9	17.7	82.3	-0.8	4.9
Growth	3.6	4.1	5.6	5.1	15.5	84.5	-0.8	8.5
Value	3.4	3.8	6.4	4.3	15.3	84.7	-0.6	8.4
Fund Characteristics								
Low Expense	2.2	3.5	5.0	6.4	23.4	76.6	-1.7	6.1
High Expense	4.6	4.3	5.9	3.1	12.0	88.0	-0.4	9.9
Low Turnover	3.3	4.0	5.1	4.8	16.0	84.0	-0.9	8.2
High Turnover	3.8	5.0	5.5	5.4	17.2	82.8	-1.2	9.2

Table IV
Cross-Sectional Distribution of the Size Coefficient

The table contains the summary statistics on the cross-sectional distribution of the size coefficient for all funds, four styles groups (small cap, large cap, growth, value), and four characteristic-sorted groups (low expense, high expense, low turnover, high turnover). It reports the first four moments, the proportions of funds with a negative and positive size coefficient, and the distribution quantiles at 10% and 90%. All cross-sectional estimates are adjusted for bias (smoothing and EIV) using our non-parametric approach.

	Moments				Proportions (%)		Quantiles (Ann.)	
	Mean (Ann.)	Volatility (Ann.)	Skewness	Kurtosis	Negative Skill	Positive Skill	10%	90%
All Funds	1.5	1.5	8.1	7.2	14.1	85.9	-0.2	3.4
Investment Styles								
Small-cap	2.0	1.5	8.2	3.7	9.1	90.9	0.1	4.1
Large-cap	1.0	1.0	7.4	2.7	17.1	82.9	-0.3	2.3
Growth	1.6	1.8	7.1	4.9	17.1	82.9	-0.5	3.9
Value	1.4	1.5	7.2	4.7	18.0	82.0	-0.4	3.3
Fund Characteristics								
Low Expense	0.9	1.3	2.9	13.2	24.4	75.6	-0.6	2.4
High Expense	1.9	2.0	7.8	4.1	14.5	85.5	-0.3	4.3
Low Turnover	1.3	1.7	3.7	8.0	19.8	80.2	-0.6	3.3
High Turnover	1.7	2.2	4.6	5.7	18.5	81.5	-0.7	4.2

Table V
Cross-Sectional Distribution of the Value Added

Panel A contains the summary statistics on the cross-sectional distribution of the lifecycle value added for all funds, four styles groups (small cap, large cap, growth, value), and four characteristic-sorted groups (low expense, high expense, low turnover, high turnover). It reports the first four moments, the proportions of funds with a negative and positive value added, and the distribution quantiles at 10% and 90%. Panel B provides the same information for the steady stated value added (as the fund reaches its average size). All cross-sectional estimates are adjusted for bias (smoothing and EIV) using our non-parametric approach.

Panel A: Lifecycle Value Added

	Moments				Proportions (%)		Quantiles (Ann.)	
	Mean (Ann.)	Volatility (Ann.)	Skewness	Kurtosis	Negative Skill	Positive Skill	10%	90%
All Funds	1.9	9.3	8.8	14.5	35.9	64.1	-2.3	11.1
Investment Styles								
Small-cap	4.1	11.6	9.8	19.9	34.8	65.2	-2.9	15.4
Large-cap	0.0	7.8	5.5	8.8	47.8	52.2	-3.1	5.2
Growth	2.3	11.3	10.5	16.8	38.0	62.0	-3.4	14.0
Value	1.5	11.9	2.1	6.1	41.5	58.5	-3.4	8.9
Fund Characteristics								
Low Expense	5.5	21.2	9.8	9.2	30.8	69.2	-2.5	31.1
High Expense	1.3	7.7	1.6	6.7	39.7	60.3	-2.9	6.9
Low Turnover	5.8	19.1	8.9	2.5	12.4	87.6	-0.2	26.4
High Turnover	1.0	13.3	4.2	7.8	44.9	55.1	-5.9	10.9

Panel B: Steady State Value Added

	Moments				Proportions (%)		Quantiles (Ann.)	
	Mean (Ann.)	Volatility (Ann.)	Skewness	Kurtosis	Negative Skill	Positive Skill	10%	90%
All Funds	7.1	15.1	9.6	21.5	27.3	72.7	-1.7	25.6
Investment Styles								
Small-cap	8.4	14.2	8.0	13.7	17.8	82.2	-1.3	24.3
Large-cap	4.9	12.1	11.7	24.4	31.7	68.3	-1.6	21.0
Growth	8.8	19.4	10.1	20.8	29.2	70.8	-2.3	33.9
Value	6.7	14.3	7.0	10.8	24.7	75.3	-1.5	25.4
Fund Characteristics								
Low Expense	10.3	30.8	8.3	16.2	32.5	67.5	-3.1	44.4
High Expense	5.5	10.3	9.0	13.5	23.0	77.0	-1.5	19.9
Low Turnover	12.9	28.5	8.5	12.9	16.5	83.5	-0.8	51.2
High Turnover	6.0	13.8	8.2	9.4	26.5	73.5	-2.1	23.0

Table VI
Optimal Versus Actual Value Added

The table compares the optimal value added with the two versions of the actual value added (lifecycle, steady state) for all funds, four styles groups (small cap, large cap, growth, value), and four characteristic-sorted groups (low expense, high expense, low turnover, high turnover). It reports the cross-sectional means for the optimal and actual value added, as well as their absolute and relative difference.

| | Mean (Ann.) | | | Abs. Difference | | Rel. Difference (%) | |
	Optimal	Value Added Lifecycle	Value Added Steady State	Value Added Lifecycle	Value Added Steady State	Value Added Lifecycle	Value Added Steady State
All Funds	13.4	1.3	9.0	12.2	4.5	9.5	66.8
Investment Styles							
Small-cap	12.7	3.1	9.1	9.6	3.6	24.2	71.5
Large-cap	12.7	-0.8	6.9	13.4	5.8	-6.2	54.3
Growth	15.0	0.9	11.0	14.1	4.1	6.1	73.0
Value	14.3	0.9	9.2	13.4	5.1	6.1	64.1
Fund Characteristics							
Low Expense	23.5	4.9	14.7	18.6	8.8	20.8	62.6
High Expense	10.4	0.6	6.9	9.8	3.5	5.6	66.6
Low Turnover	27.3	4.8	16.7	22.5	10.6	17.7	61.3
High Turnover	14.2	-0.9	7.6	15.1	6.6	-6.3	53.7

Table VII
Fund Fees and Size

The table contains the summary statistics on the cross-sectional distribution of fund fees and size for all funds, four styles groups (small cap, large cap, growth, value), and four characteristic-sorted groups (low expense, high expense, low turnover, high turnover). It reports the mean, volatility, and the distribution quantiles at 10% and 90%.

	Mean		Volatility		Quantile 10%		Quantile 90%	
	Fees	Size	Fees	Size	Fees	Size	Fees	Size
All Funds	1.25	753	0.41	2076	0.79	43	1.80	1645
Investment Styles								
Small-cap	1.37	410	0.36	673	0.97	48	1.87	907
Large-cap	1.16	1068	0.39	2953	0.73	45	1.71	2298
Growth	1.30	769	0.41	2059	0.87	47	1.85	1682
Value	1.18	980	0.38	2690	0.74	47	1.66	1851
Fund Characteristics								
Low Expense	0.82	1411	0.18	3563	0.56	63	1.01	2958
High Expense	1.73	344	0.28	727	1.44	36	2.10	785
Low Turnover	1.14	1475	0.37	3980	0.72	59	1.58	3246
High Turnover	1.30	650	0.37	1272	0.89	56	1.79	1457

Table VIII
Cross-Sectional Distribution of Gross Alpha

The table contains the summary statistics on the cross-sectional distribution of the gross alpha for all funds, four styles groups (small-cap, large-cap, growth, value), and four characteristic-sorted groups (low-expense, high-expense, low-turnover, high turnover). It reports the first four moments, the proportions of funds with a negative and positive gross alpha, and the distribution quantiles at 10% and 90%. All cross-sectional estimates are adjusted for bias (smoothing and EIV) using our non-parametric approach.

	Moments				Proportions (%)		Quantiles (Ann.)	
	Mean (Ann.)	Volatility (Ann.)	Skewness	Kurtosis	Negative Skill	Positive Skill	10%	90%
All Funds	0.8	1.3	3.2	1.2	27.8	72.2	-0.9	2.7
Investment Styles								
Small-cap	1.6	1.2	6.4	-1.9	12.3	87.7	-0.4	3.7
Large-cap	0.3	0.9	1.6	3.1	36.0	64.0	-0.9	1.7
Growth	0.8	1.5	2.6	1.7	31.6	68.4	-1.3	2.9
Value	1.0	1.6	4.2	2.1	26.5	73.5	-1.0	3.0
Fund Characteristics								
Low Expense	0.4	1.4	2.4	2.7	38.8	61.2	-1.3	2.0
High Expense	1.1	1.8	2.9	3.3	25.2	74.8	-1.2	3.6
Low Turnover	0.9	1.5	2.8	0.8	27.8	72.2	-1.1	3.0
High Turnover	0.7	1.9	2.8	3.1	34.8	65.2	-1.6	3.1

Table IX
Cross-Sectional Distribution of Net Alpha

The table contains the summary statistics on the cross-sectional distribution of the net alpha for all funds, four styles groups (small-cap, large-cap, growth, value), and four characteristic-sorted groups (low-expense, high-expense, low-turnover, high turnover). It reports the first four moments, the proportions of funds with a negative and positive net alpha, and the distribution quantiles at 10% and 90%. All cross-sectional estimates are adjusted for bias (smoothing and EIV) using our non-parametric approach.

	Moments				Proportions (%)		Quantiles (Ann.)	
	Mean (Ann.)	Volatility (Ann.)	Skewness	Kurtosis	Negative Skill	Positive Skill	10%	90%
All Funds	-0.5	1.3	-2.8	2.2	63.5	36.5	-2.2	1.4
Investment Styles								
Small-cap	0.2	1.3	-0.6	0.8	46.1	53.9	-1.8	2.4
Large-cap	-0.8	0.9	-5.3	1.6	79.1	20.9	-2.0	0.5
Growth	-0.5	1.6	-2.8	2.8	64.3	35.7	-2.6	1.6
Value	-0.2	1.5	0.1	2.0	56.0	44.0	-2.1	1.7
Fund Characteristics								
Low Expense	-0.4	1.3	-1.0	2.4	63.4	36.6	-2.1	1.2
High Expense	-0.6	1.9	-2.5	3.8	61.1	38.9	-3.0	2.0
Low Turnover	-0.3	1.5	-1.7	1.6	56.5	43.5	-2.2	1.8
High Turnover	-0.6	2.0	-1.3	3.3	62.8	37.2	-2.8	1.8

Figure 1
Bias for the First Dollar Alpha

This figure plots the two bias components (smoothing and Error-in-Variable (EIV)) for the cross-sectional distribution of the first dollar (fd) alpha. Each component is obtained from the normal reference model whose five parameters are estimated using the estimated gross alpha and its asymptotic variance for all funds in the population (n=2,265). The estimated values for the mean and volatility of the *true* fd alpha are equal to 0.28% and 0.42% per month. The mean and volatility of the log of the asymptotic variance are equal to -6.0 and 1.1. Finally, the correlation coefficient is equal to 0.27.
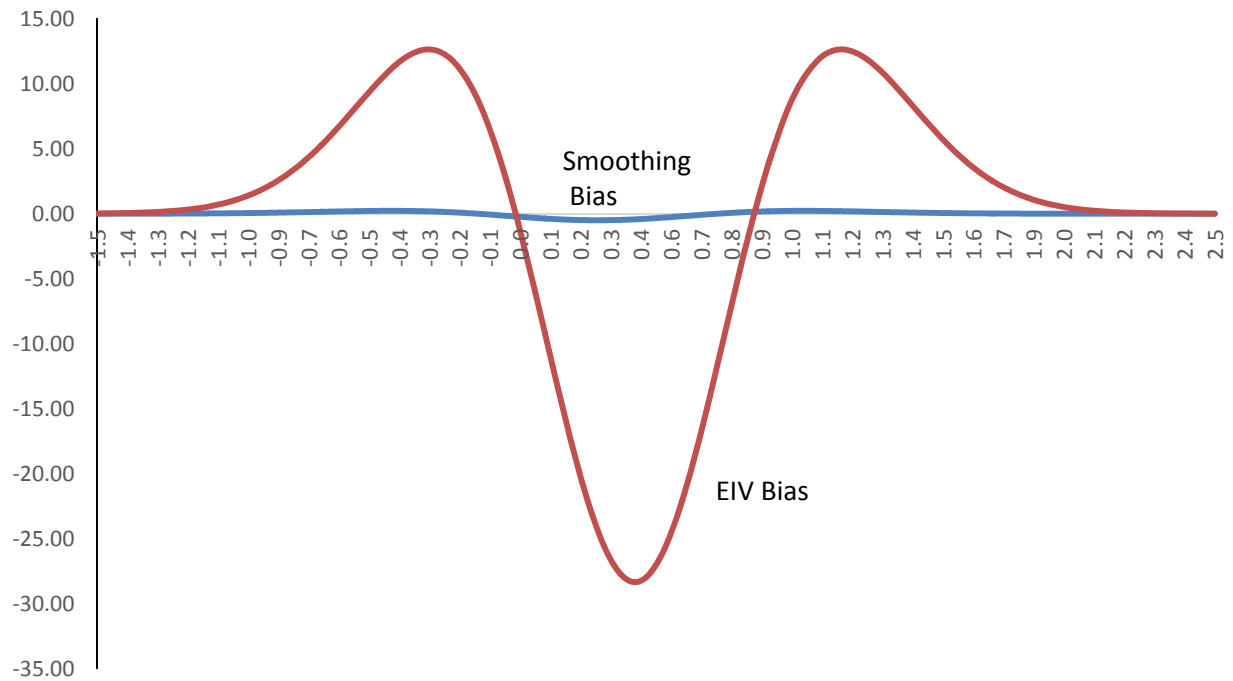
Figure 2
Cross-sectional Distribution of the First Dollar Alpha:
Analysis across Fund Groups

Panel A plots the cross-sectional densities of the first dollar alpha for small cap and large cap funds. Panel B compares growth and value funds. Panel C compares low expense and high expense funds. Finally, Panel D compares low turnover and high turnover funds. All the estimated densities are adjusted for bias (smoothing and EIV) using our non-parametric approach.
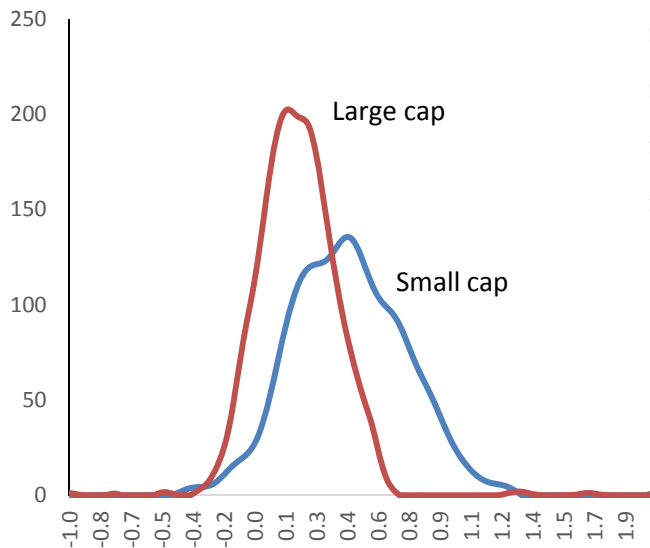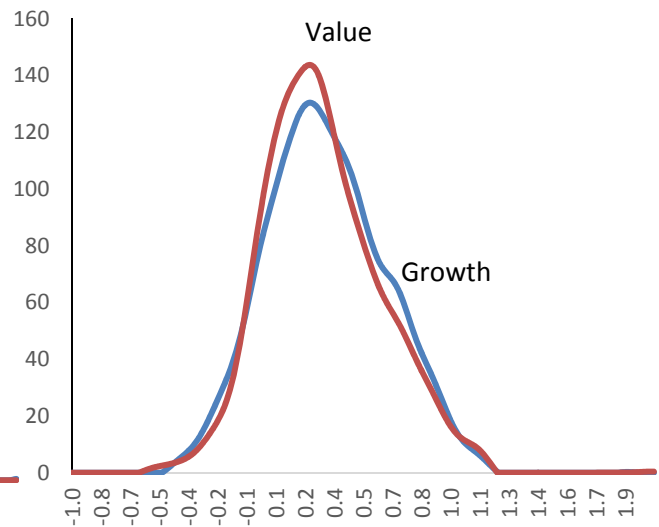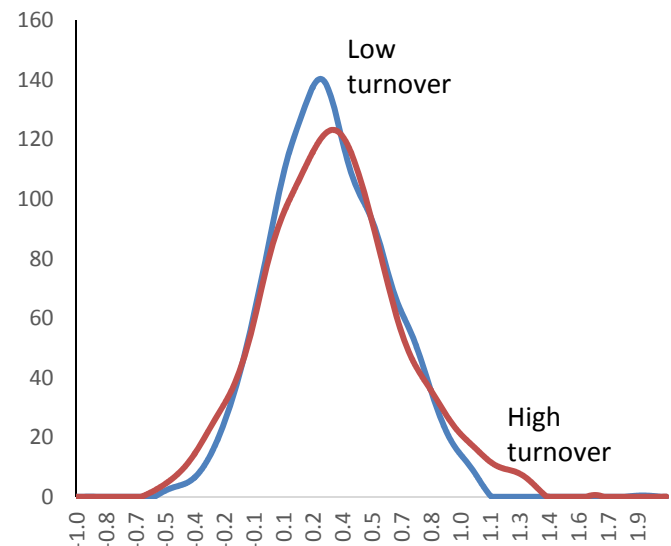
Panel A: size groups

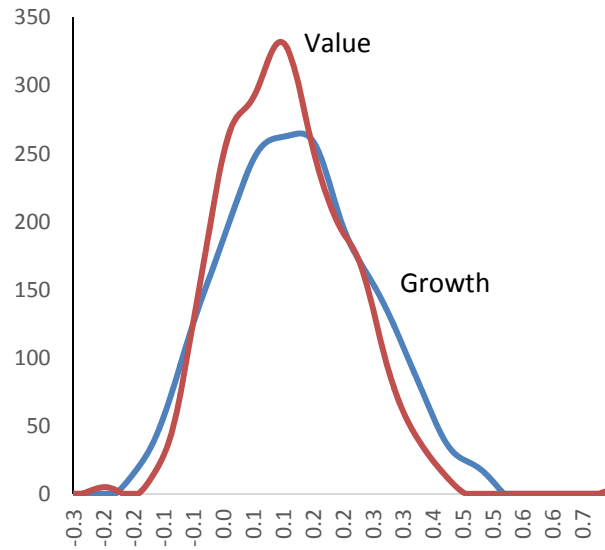Panel B: Value/growth groups
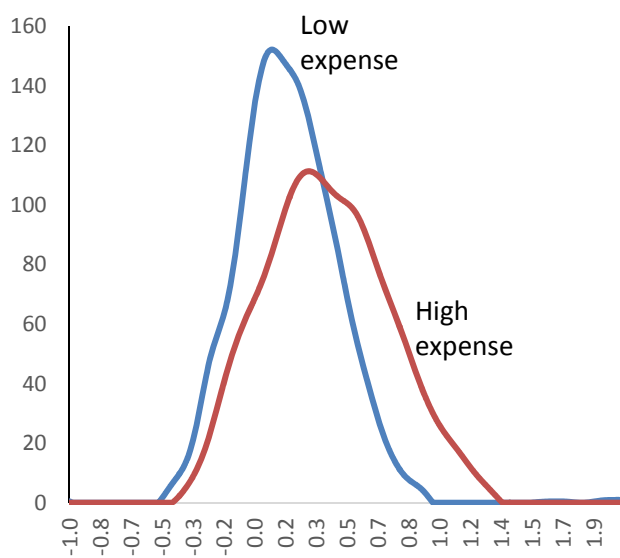
Panel C: Expense groups

Panel D: Turnover groups

Figure 3
Cross-sectional Distribution of the Size Coefficient:
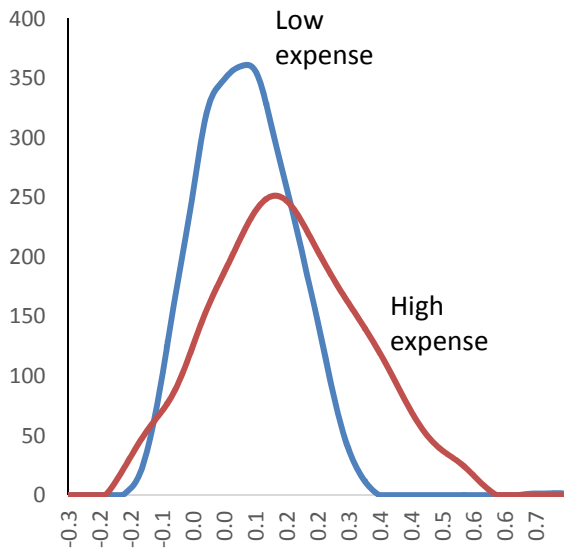## Analysis across Fund Groups

Panel A plots the cross-sectional densities of the size coefficient for small cap and large cap funds. Panel B compares growth and value funds. Panel C compares low expense and high expense funds. Finally, Panel D compares low turnover and high turnover funds. All the estimated densities are adjusted for bias (smoothing and EIV) using our non-parametric approach..
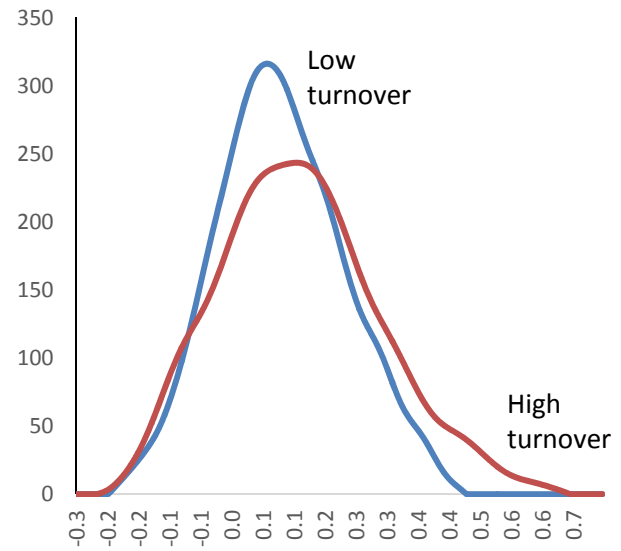

Panel A: size groups

Panel B: Value/growth groups

Panel C: Expense groups

Panel D: Turnover groups